THE DETERMINATION OF
EMPIRICAL AND ANALYTICAL SPACECRAFT PARAMETRIC CURVES
- THEORY AND METHODS -

PROGRESS REPORT II

Submitted

by the

TEXAS A&M RESEARCH FOUNDATION

to the

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

for NASA Grant

SC-NGR-44-001-027

Prepared

by

THE INDUSTRIAL ENGINEERING DEPARTMENT

of

TEXAS A&M UNIVERSITY

May 1, 1966

TABLE OF CONTENTS

FOREWORD

This document represents the second progress report on the NASA research grant NGR 44-001-027. The report is divided into eight parts. A summary of these parts is presented below in order to indicate the general nature of each section.

Part I is a report on the research outlined in Part III of the first progress report. In general, it deals with the problem of fitting a nonlinear function with constraints on the parameters. Numerical examples are presented in order to clarify parts of the development so that application of the principles can be made.

Part II is a most significant development in the areaof constructing mathematical models with a limited amount of data. It considers the form of the function selected to represent the general response of the dependent variable over the applicable range of the independent variable. By restricting the first order terms of the variables to the preselected function, it is possible to determine a curve with more parameters than data points. By repeating this process for each higher order term, it is possible to expand a single equation to a large number of terms, using the least squares criteria for parameter selection.

Part III is a development of a computer program for use in evaluation of the cost estimating relationship developed in Part II of this report. The program was used to checkout and evaluate the overall consistency of the cost equation. By utilizing three dimensional co-ordinate paper to plot the results, it was possible to judge the behavior of the CER over a rather extended range. The results produced by this program tend to support the approach developed in Part II.

Part IV is a report on the development of a dynamic programming algorithm for solution of problems that were discussed with another NASA cost research contractor. This is an applied research problem which provided the opportunity to introduce a new application of dynamic programming as well as contribute to the technology in the cost research area. The basic problem was the combination of individual predictors of the same cost by weighting each term so as to minimize the error sum of squares.

Part V is the development of a methodology for the determination of run-out costs for partially completed subsystem. The particular approach was to use the Gemini data in a general manner in order to predict the run-out of Apollo subsystems. This study produced a number of tangential investigations as well as a computerized algorithm for determining run-out costs. The applied research associated with this problem was essentially initiated and completed within this reporting period. This was in response to a specific problem suggested by NASA/MSC/LRP.

Part VI is a discussion of the present status of the research which is being conducted in the area of quantification of expertise. This part of the report provides a brief history of the area and reports on current progress and questionnaires developed for initial investigations.

Part VII is of the same general nature as Part VI of the first progress report in that it represents research in an area related to general cost models. More specifically, it develops a concept for sophisticated production cost models which consider the recycle, cleanup, closed loop and recycle with a primary loss cases.

Part VIII is the informal consulting memoranda generated during the time period covered by this report.

PART I

CONSTRAINED ESTIMATION IN NONLINEAR MODELS: A MATHEMATICAL PROGRAMMING APPROACH.

By

William P. Cooke

For

Industrial Engineering Department

Texas A&M University

May 1966

INTRODUCTION

The problem of "estimation of parameters with constraints for models which are nonlinear in the parameters" was discussed in Part III of Progress Report I for NASA Grant SC-NGR-44-001-027, December, 1965. In that report the general problem was presented, the possible approaches to its solution were discussed, a bibliography was included, and a connection was made of this pure research problem with the specific problems anticipated in the construction of cost models.

This section of this report will present some of the theoretical results obtained, a discussion of a promising method of attack on the estimation problem, and a specific example in which a solution was obtained.

Only the estimation problem has been investigated in detail; the question of properties of the estimates being deferred until the general estimation theory has been adequately developed. It is clear, however, that the estimates should be either "best linear estimates" (where "best" means minimum mean-square error) or estimates which depart from that criterion only slightly. In fact, the method of estimation has been formulated in a manner designed to insure that property or some extension of it.

Therefore, while estimation only is discussed in this report, the question of properties of the estimators has actually been considered concurrently.

THE GENERAL CONSTRAINED NONLINEAR PROBLEM

The objective is to estimate the parameter vector Θ in the function $y = f\ (X; \Theta)$, where $X = (x_1, \ldots, x_k)$ is a vector of observable variables, y is observable (the dependent variable), $\Theta = (\theta_1, \ldots, \theta_p)$ is a vector of fixed but unknown parameters, and Θ is restricted (constrained) to a certain region of E^p (p - dimensional Euclidean space).

The problem as stated above is the most general of estimation problems. No specification of the function f or of the nature of the constraints on Θ is made. In fact, all problems of parameter estimation (based on sample observations) for which there exist standard methods of solution are simply special cases of this general problem.

For example, if $f\ (X, \Theta)$ is linear in the parameters θ_i and Θ is unconstrained, we have the standard multiple regression problem. If $f\ (X, \Theta)$ is linear in Θ and Θ is constrained by linear inequalities while each element of Θ is required to be positive, the estimation problem can be formulated as a quadratic programming problem.

The case were $f\ (X, \Theta)$ is nonlinear in Θ but Θ is unrestricted has been extensively investigated and various iterative techniques have been proposed for its solution. As with any iterative technique however, the choice of an initial vector Θ_I is critical. This same difficulty has arisen in the problem where, in addition to $f\ (X, \Theta)$ being nonlinear, Θ is constrained. This last problem is the main topic to be discussed, but in the discussion of its solution we also are able to propose a possible criterion for an initial Θ_I in the unrestricted case.

Before proceeding to the theoretical basis for solution to the problem, let us first consider the problem in somewhat less general language so that the reader with a limited mathematical background may get a "feel" for the problem.

As a vehicle for discussion, let us formulate a simple cost estimation problem. Suppose that it is desired to estimate the cost of a certain spacecraft subsystem and we make the assumption that some mathematical function of the two variables x_1 (subsystem weight) and x_2 (number of flight functions) adequately describes cost. We might go a step further and say that we know the form of the function is $y = x_1^{\theta_1} + x_2^{\theta_2}$, where y represents cost, and θ_1 and θ_2 are constants (parameters). Then, if θ_1 and θ_2 are known, we may simply "plug in" the values of x_1 and x_2 in order to know the cost y exactly for that pair of values x_1 and x_2. In other words, if θ_1 and θ_2 are known, there is no problem. We do not estimate costs, we <u>know</u> them.

Unfortunately one is rarely in this ideal situation. In fact, we consider ourselves quite fortunate to be able to specify the <u>form</u> of the function. Suppose, however, that we can, but that we do not know θ_1 and θ_2. Suppose further that our experience tells us that both θ_1 and θ_2 must be positive constants. We now have a constrained estimation problem, for what we would like to do is the following.

We have a little bit of knowledge about the parameters θ_1 and θ_2 (they are positive). We also have some historical data pertaining to costs of subsystems of this type. That is, for certain pairs (x_1, x_2) we have observed the actual cost y for this subsystem. For example, when weight (x_1) was 9,000 pounds and number of functions (x_2) was 5, cost (y), was known. Obviously if our functional form is correct, and if we know θ_1 and θ_2, then the actual observed y would have been equal to $(9{,}000)^{\theta_1} + (5)^{\theta_2}$.

We now propose to use this data as shown below. Suppose we have three such sets of data. We form the following three equations:

$$y_1 = x_{11}^{\theta_1} + x_{21}^{\theta_2}$$
$$y_2 = x_{12}^{\theta_1} + x_{22}^{\theta_2}$$
$$y_3 = c_{13}^{\theta_1} + x_{23}^{\theta_2} ,$$

where (x_{11}, x_{21}) represents the known x_1 and x_2 for observed cost y_1, (x_{12}, x_{22}) are associated with y_2, etc. We ask ourselves the question: "What two positive numbers θ_1 and θ_2 satisfy these three equations?". If there are two such numbers, they might be determined by inspection, or "trial and error", or by some "sophisticated" mathematical technique. In fact we are certain that eventually, by some means, we will find those two positive numbers.

Usually though, the situation is not quite so clearly defined. For one thing, the form of the function might be only an approximation to the actual cost function (in fact, it usually is an approximation). Then there will not exist any pair of numbers θ_1 and θ_2 that satisfy the three equations exactly, and a "trial and error" method would have us "trying and erring" indefinitely.

Another, though not so severe, difficulty is that the x's or the y might not be exactly known. Usually the x's will be known exactly for all practical purposes, but y itself might only be an estimate or, more likely, impossible to determine exactly (what is the exact cost of solder-joint number 2437?).

After considering the above points we conclude that it is virtually impossible to ever really determine θ_1 and θ_2, so the best we can hope to do is to estimate them in such a way that the estimated value is expected to be very near the actual value, which in turn will enable us to estimate costs adequately. We are now at the heart of the problem.

There are standard statistical techniques which enables us to estimate parameters and say something about how good these estimates are if the problem is linear, that is, if y has a form like $y = \theta_1 x_1 + \theta_2 x_2$ or $y = \theta_1 + \theta_2 x_1 x_2$ instead of being nonlinear, as $y = x_1^{\theta_1} + x_2^{\theta_2}$ ("linear" and "nonlinear" refer to the θ's, not the x's). Even if the problem is nonlinear, it is usually approached by first approximating y by a linear function and then using a response-surface technique on this linear approximation (this is not always satisfactory).

However, the problem of the constraints now enters. Suppose that θ_1 and θ_2 are very nearly zero, for example, that their actual values are $\theta_1 = .002$ and $\theta_2 = .001$. Suppose further that because of the fact that we only have observed a few points, and at least one observation is somewhat inaccurate, our standard procedure tells us that the estimate of θ_1 is $-.0003$ and of θ_2 is .0012. Immediately this is an unuseable result, for we know that the actual θ_1 must be positive. One could in fact say that $\theta_1 = 0$ is a better estimate than the negative one we obtained. But is $\theta_1 = 0$ the best estimate?

The above discussion of an elementary constrained estimation problem was intended to point out at least a few of the reasons why the general problem is being considered. It should also establish a relationship between the purely academic research being carried out under the grant with the specific problem of cost estimation. If a nonlinear cost equation is specified, and if some information is available about the location (or size) of the constants in the equation, then the estimation problem is a special case of the general problem we now attempt to solve.

FORMULATION OF THE ESTIMATION PROBLEM AS A MATHEMATICAL PROGRAMMING PROBLEM

In the general problem originally proposed, we wish to estimate the parameter vector θ in the function $y = f(X, \theta)$ when θ is constrained to some region of Euclidean space. Suppose for the moment we regard θ as unrestricted and consider only the unconstrained problem.

Let us make the assumption that while we cannot observe y exactly, we are able to observe

$$y_i = f(X_i; \theta) + e_i ,$$

where y_i is the observed response corresponding to an input vector

$X_i = (x_{1i}, x_{2i}, \ldots, x_{ki})$ and e_i is a random error whose expectation is zero and whose fixed but unknown variance is σ^2.

We now consider the set of observed responses

$$y_i = f\ (X_i;\ \Theta) + e_i\ ,\ i = 1,\ 2,\ \ldots,\ n\ ,$$

a set of n data-points, or n observations. This set can be described compacty in matrix notation as

$$Y = U\ (X;\ \Theta) + e\ ,$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\quad U\ (X;\ \Theta) = \begin{bmatrix} f\ (X_1;\ \Theta) \\ f\ (X_2;\ \Theta) \\ \vdots \\ f\ (X_n;\ \Theta) \end{bmatrix},\quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

We further assume that the e_i's are independent with common variance σ^2.

A standard statistical procedure for estimating Θ is least-squares; that is, find that vector Θ which minimizes the sum of squares of the "residuals", or e_i's. Under the assumptions stated, such solution is a best linear unbiased estimator for Θ, in the sense of the well-known Gauss-Markoff Theorem.

In matrix terms, the problem is to minimize

$$e'e = [Y - U\ (X;\ \Theta)]'[Y - U\ (X;\ \Theta)]\quad ,$$

which is a function only of the p-vector Θ, since all the y's and x's are observed (known) values.

This problem has a very simple solution if $f\ (X;\ \Theta)$ is linear in the Θ's. When that is the case, e'e is a quadratic form in Θ, which, when each partial derivative of e'e is set equal to zero, yields a system of simultaneous linear equations in Θ. (Recall that such a system is a necessary condition for the existence of a minimum).

Now, if $f(\underset{\cdot}{X};\ \Theta)$ is not linear in Θ, then e'e will not be a simple quadratic form in Θ, and our system of equations representing the necessary condition for a minimum will usually be quite difficult, if not impossible to solve analytically.

A numerical solution of such a system is possible, but usually not practicable.

One resolution of this difficulty, which is a well known and well documented approach, is the "linearization" of $f(X; \Theta)$ using a first-order Taylor's Series expansion about some particular Θ, say Θ^*, and then minimizing $e'e$, where $U(X; \Theta)$ is replaced by $U^*(X; \Theta)$ based on the approximation. The well known method of "steepest ascent", an iterative technique, is ordinarily employed to accomplish the minimization. It is known to converge to a solution if the initial vector Θ_I, to begin the first iteration, is in the neighborhood of the actual solution. Obviously some criterion is required to assure a good Θ_I, and various conditions on and modifications of the technique have been proposed. A good discussion of these modification is found in references [1] and [2].

Of course a Taylor's Series approximation is not the only method of approach to solving a problem of this type. Another procedure, which is known to converge, has been given by Dr. H. O. Hartley in reference [3]. Our suggested approach in this report will employ the Taylor's Series approximation technique, but in subsequent research we will explore the possibilities of using Dr. Hartley's procedure as applied to the constrained estimation problem.

The problem of constraints on Θ is not considered in the above discussion. If Θ is restricted, then the approach we suggest is a modification of the Taylor's Series approach to the unconstrained problem, with a mathematical programming rather than a "steepest ascent" method for the solution. The theory follows.

Consider now the point $\Theta^{*\prime} = [\Theta_1^*, \ldots, \Theta_p^*]$ in the (constrained) Θ- space. We shall write a Taylor's expansion of $f(X; \Theta)$ about this point, neglecting terms of order 2 and higher.

Form the equation

$$f(X_i; \Theta) \doteq f(X_i; \Theta^*) + \sum_{j=1}^{p} (\Theta_j - \Theta_j^*)\, f_j(X_i; \Theta^*) ,$$

where

$$f_j(X_i;\ \theta^*) = \left.\frac{\partial f\,(X_i;\ \theta)}{\partial \theta_j}\right|_{\theta = \theta^*} .$$

Then

$$U\ (X\ ;\ \theta) = \begin{bmatrix} f\ (X_1\ ;\ \theta) \\ f\ (X_2\ ;\ \theta) \\ \vdots \\ f\ (X_n\ ;\ \theta) \end{bmatrix} \doteq \begin{bmatrix} f\ (X_1;\ \theta^*) & f_1(X_1;\ \theta^*) & \ldots f_p\ (X_1;\ \theta^*) \\ f\ (X_2;\ \theta^*) & f_2(X_2;\ \theta^*) & \ldots f_p\ (X_2;\ \theta^*) \\ \vdots & \vdots & \vdots \\ f\ (X_n;\ \theta^*) & f_1(X_n;\ \theta^*) & \ldots f_p\ (X_n;\ \theta^*) \end{bmatrix} \begin{bmatrix} 1 \\ \theta_1 - \theta_1{}^* \\ \vdots \\ \theta_p - \theta_p{}^* \end{bmatrix} .$$

or we have approximated $U\ (X;\ \theta)$.

Let

$$A = \begin{bmatrix} f\ (X_1;\ \theta^*) \\ f\ (X_2;\ \theta^*) \\ \vdots \\ f\ (X_n;\ \theta^*) \end{bmatrix} ,\ B = \begin{bmatrix} f_1(X_1;\ \theta^*) & f_2(X_1;\ \theta^*)\ldots & f_p\ (X_1;\ \theta^*) \\ f_1(X_2;\ \theta^*) & f_2(X_2;\ \theta^*)\ldots & f_p\ (X_2;\ \theta^*) \\ \vdots & \vdots & \vdots \\ f_1(X_n;\ \theta^*) & f_2(X_n;\ \theta^*)\ldots & f_p\ (X_n;\ \theta^*) \end{bmatrix} ,$$

$$\theta - \theta^* = \begin{bmatrix} \theta_1 - \theta_1^* \\ \theta_2 - \theta_2^* \\ \vdots \\ \theta_p - \theta_p^* \end{bmatrix} ,\ \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} ,\quad \theta^* = \begin{bmatrix} \theta_1^* \\ \theta_2^* \\ \vdots \\ \theta_p^* \end{bmatrix} .$$

Notice that all the elements of A and B are known, depending only on the known X's and the chosen θ^*,

Then $U\ (X;\ \theta) \doteq A + B\ (\theta - \theta^*)$, and we wish to minimize

$$\begin{aligned} e'e &\doteq [Y - A - B\theta + B\theta^*]\ '\ [Y - A - B\theta + B\theta^*] \\ &= [(Y - A + B\theta^*) - B\theta]\ '\ [(Y - A + B\theta^*) - B\theta] \\ &= [(Y' - A' + \theta^{*\prime}B') - \theta'B']\quad [(Y - A + B\theta^*) - B\theta] \\ &= (Y' - A' + \theta^{*\prime}B')\ (Y - A + B\theta^*) - (Y' - A'\theta^{*\prime}B')\ (B\theta) \\ &\quad - \theta'B\ (Y - A + B\ \theta^*) + \theta'B'B\ \theta \\ &= K - [2(Y' - A' + \theta^{*\prime}B')\ B\theta - \theta'B'B\theta] \quad , \end{aligned}$$

where $K = (Y' - A' + \theta^{*\prime}\ B')\ (Y - A + B\theta^*)$ is a known constant.

Then minimizing e'e is equivalent to <u>maximizing</u> z, where

$$z = [2\ (Y' - A' + \theta^{*\prime}B')\ B]\theta + \theta'(-B'B)\ \theta.$$

If there are no constraints on θ, then this is the usual (references [1] and [2]) Taylor's series approach, except the function z has been written as a quadratic form in θ, while usually a quadratic form in $(\theta - \theta^*)$ is used in the references [1] and [2]. Our form of z is important from a mathematical programming viewpoint, however. Note that we could perhaps simplify the equation slightly by maximizing z/2 rather than z, but then the matrix 1/2 B'B rather than B'B would appear, and it was decided that the problem might be better handled if the form B'B is used.

Let us now state what can be done, from a mathematical programming standpoint, with this problem if Θ is constrained.

If B'B is positive definite, $z = z(\Theta)$ is a concave function. Suppose the constraints on Θ are of the form

$$g_r(\Theta) \leq 0 \ , \ r = 1, 2, \ldots, m$$

$$\theta_i \geq 0 \ , \ i = 1, 2, \ldots, p$$

and that $g_r(\Theta)$ is a convex function of θ for all r.

There is an algorithm available for the solution of the problem: maximize $z(\Theta)$, where $z(\Theta)$ is a concave function, subject ot the conditions

$$g_r(\Theta) \leq 0 \ , \ r = 1, 2, \ldots, m$$

and

$$\theta_i \geq 0 \ , \ i = 1, 2, \ldots, p$$

where the $g_r(\Theta)$ are convex functions. The algorithm is described in reference [4] and offers a complete solution for this particular case. It should be noted that the restriction on Θ is of a type that is rather common; that is, very often the constraints are convex functions in Θ, so that a rather broad class of problems may be solved using the algorithm.

The more restrictive requirement, for the Taylor's series approach, is that $z(\Theta)$ be a concave function (equivalently, that B'B be positive definite). Recall that B is determined by the choice of θ^*. While θ^* was chosen so that it satisfied the original constraints on Θ, we are <u>not</u> assurred that this choice will force B'B to be positive definite; and in fact it does not.

We now propose an approach to the solution of the problem which we believe has much potential value. We propose simply that we utilize the requirement of the algorithm, that z (θ) be concave, as an additional restriction; not on θ, but on our choice of θ^*. In other words, we force z (θ) to be concave by choosing as an initial vector a θ^* such that $B'B$ is positive definite. Equivalently, we have constructed a function z (θ), which we must maximize, and we simply require it to have a form which possess a unique maximum; certainly a reasonable requirement.

Then, if the constraints on θ are convex, the algorithm of [4] may be used for a first iteration (a solution to the approximate problem). This yields a first estimate $\hat{\theta}$, which then becomes the choice of θ^* for a second iteration, and so on.

As one can see, a rather large class of problems may be solved using this procedure. The method is at present only in its first stages of development, as it remains to prove that it will always converge to a solution of the original problem, or else to find a set of conditions that will guarantee convergence. Certainly, in the unrestricted case, such conditions are available (references [1] and [2]), leading us to believe that they also exist in the restricted case.

The Hartley-Hocking algorithm (reference [4]) has been recently programmed for computer by L. Claypool of the Institute of Statistics, Texas A&M University, and the program is now available for our use.

AN EXAMPLE PROBLEM

We will now demonstrate a problem of this type for which a solution was obtained. Since the use of the Hartley-Hocking algorithm is only feasible on the computer, the example is stated in such a way that a solution may be obtained without the algorithm, while still illustrating the utility of the Taylor approximation and the requirement of positive definiteness. Actually, the problem is somewhat more restrictive in that an equality constraint rather than an inequality was applied to Θ. From a mathematical programming viewpoint, an equality is more restrictive than an inequality, but from a classical viewpoint it simplifies the problem somewhat. We expect, of course, that the cost estimation problems will involve inequalities.

Let us now attempt to estimate the parameters for the function

$$y_i = \theta_1 \theta_2^{\ x_i} + \theta_3 \theta_4^{\ x_i} \ ,$$

subject to the conditions

$$\sum_{j=1}^{4} \theta_j = 3 \ ,$$

and

$$\theta_j \geq 0 \ , \ i = 1, 2, 3, 4 \ ,$$

where we have the following observed data:

i	x_i	y_i
1	0	1.25
2	1	1.00
3	2	.97
4	3	.96
5	4	1.05

Then

$$f \ (x_i; \Theta) = \theta_1 \theta_2^{\ x_i} + \theta_3 \theta_4^{\ x_i}$$

$$f_1 \ (x_i; \Theta) = \theta_2^{\ x_i}$$

$$f_2 \ (x_i; \Theta) = x_i \theta_1 \theta_2^{\ x_i - 1}$$

$$f_3 \ (x_i; \Theta) = \theta_4^{\ x_i}$$

$$f_4 \ (x_i; \Theta) = x_i \theta_3 \theta_4^{\ x_i - 1} \ .$$

First of all we ignore the constraints and find an "intuitive" solution (this was done by observing the function and making the conjecture that data point number 1 might be the most accurate observation and that point number 5 would be most "sensitive"). What we are attempting to do, of course, is to arrive at a feasible Θ^*, our initial vector. Actually we simply look for a Θ^* which satisfies all the conditions we impose and which does us the favor that the sum of squares of deviations is small, suspecting that the θ for which the sum of squares is minimum will be in the neighborhood of θ^*. These considerations lead to

$$\tilde{\theta} = \begin{bmatrix} .53 \\ 1.18 \\ .72 \\ .44 \end{bmatrix} .$$

Adjusting $\tilde{\theta}$ slightly, we arbitrarily decide to choose θ^* so that

$$\theta^* = \begin{bmatrix} .56 \\ 1.21 \\ .76 \\ .47 \end{bmatrix} .$$

Observe that $\sum_{j=1}^{4} \theta^*_j = 3$, and that $\theta_j^* \geq 0$, so that the constraints on θ are satisfied by θ^*. A simple calculation assures us that B'B is positive definite, so we are ready to begin the first iteration.

In the notation of the general problem, we find

$$A' = (1.32,\ 1.03,\ .99,\ 1.07,\ 1.24)$$

$$\theta^{*\prime} = (.56,\ 1.21,\ .76,\ .47)$$

$$B'B = \begin{bmatrix} 12.31 & 15.51 & 2.17 & 3.53 \\ 15.51 & 23.98 & 1.01 & 3.89 \\ 2.17 & 1.01 & 1.28 & .58 \\ 3.53 & 3.89 & .58 & 1.43 \end{bmatrix}$$

$$2[Y' - A' + \theta^{*\prime} B']B = (56.46,\ 78.44,\ 7.15,\ 15.29) .$$

We may now form our function $z(\theta)$, which is written out in full below.

$$\begin{aligned} z(\theta) = {} & 56.46\theta_1 + 78.44\theta_2 + 7.15\theta_3 + 15.29\theta_4 \\ & - 12.31\theta_1^2 - 23.98\theta_2^2 - 1.28\theta_3^2 - 1.43\theta_4^2 \\ & - 31.02\theta_1\theta_2 - 4.34\theta_1\theta_3 - 7.06\theta_1\theta_4 \\ & - 2.02\theta_2\theta_3 - 7.78\theta_2\theta_4 - 1.16\theta_3\theta_4 . \end{aligned}$$

Since we have the equality constraint $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 3$, and θ^* was definitely composed of positive elements, we simply ignore the constraint $\theta_j \geq 0$ and attack the problem as a classical Lagrange - multiplier problem. If the solution satisfies $\theta_j \geq 0$, then the constraint is "inactive". If, however, it does not, then we may reconsider the problem either by the Hartley-Hocking algorithm or as a quadratic programming problem. In any event, a method exists for solving the first iteration.

Fortunately, from the standpoint of hand calculations, the Lagrange-multiplier approach works, yielding the vector $\hat{\theta}' = [.53, 1.17, .72, .58]$ which is now our estimate of θ'.

Recall now that the vector A exhibits the values of the original function $f(X_i; \theta)$ if a particular θ is used, let A^* denote the value of A when $\theta = \theta^*$, let $\hat{A}$ denote the value of A when $\theta = \hat{\theta}$, and recall that Y is the vector of observed y's.

Then

$$A^{*\prime} = [1.32, 1.03, .99, 1.07, 1.24]$$

$$\hat{A}' = [1.25, 1.04, .98, .99, 1.07] \quad ,$$

while

$$Y' = [1.25, 1.00, .97, .96, 1.05] \quad .$$

It is apparent that $\sum_{i=1}^{5} (y_i - a_i^*)^2 > \sum_{i=1}^{5} (y_i - \hat{a}_i)^2$, or $\hat{\theta}$ given an improvement in the sum of squares when compared with θ^*.

Specifically,

$$\sum_{i=1}^{5} (y_i - a_i^*)^2 = .0544$$

while

$$\sum_{i=1}^{5} (y_i - \hat{a}_i)^2 = .0030 \quad ,$$

a significant decrease in the sum of squares.

Notice that <u>we are no longer talking about the approximating function</u>, but are referring to the sum of squares of deviations of the observed from the calculated y's when in $\theta_1\theta_2^{x_i} + \theta_3\theta_4^{x_i}$ we let $\theta_1 = \theta_1^*$, $\theta_2 = \theta_2^*$, $\theta_3 = \theta_3^*$, $\theta_4 = \theta_4^*$, compared with evaluating the same function when $\theta_1 = \hat{\theta}_1, \theta_2 = \hat{\theta}_2 . \theta_3 = \hat{\theta}_3 . \theta_4 = \hat{\theta}_4$. In other words, those sums of squares are in terms of the <u>original</u> problem; minimize

$$\sum_{i=1}^{5} [y_i - \theta_1 \theta_2^{x_i} - \theta_3 \theta_4^{x_i}]^2 \quad .$$

Then certainly the use of the approximation has resulted in a significant decrease in the sum of squares we originally considered. We anticipate that if we now begin the problem chosing $\hat{\theta}$ as the point about which we make the Taylor's expansion, that a decrease in the sum of squares will again be obtained. Hence, we could either decide that .003 was acceptable as essentially a zero sum of squares, or else continue with the process until either zero or else no further improvement is obtained.

As a point of interest, the same problem, with the same function and the same data, was considered when the linear constraint $\sum_{j=1}^{4} \theta_j = 3$ was replaced by the nonlinear (but convex) constraint $\sum_{j=1}^{4} \theta_j^2 = 2.5$. The question was raised as to what effect would be obtained if a Taylor approximation was applied to the constraint as well as to the objective function. It was found that the same order of magnitude of reduction of the sum of squares was obtained! The purpose of that investigation was to attempt to formulate the problem as a special type known as a quadratic programming problem, rather than using the more general convex programming approach. This leads to some even more interesting conjectures, but it is felt that convex programming is the more fruitful avenue of approach to the general problem.

PROPOSED FUTURE RESEARCH: CONCLUSION

The areas of subsequent research on the general problem will be:

(1) The investigation of the "restrictiveness" of the requirement that B'B be positive definite.

(2) Conditions for convergence of the iterations to the true solution.

(3) The possibility that the requirement of the positive definiteness of B'B be used as an additional set of constraints on θ as well as θ^*. (This is currently being applied to a different but related problem in response-surface techniques and appears to be very promising there).

(4) The modification of Hartley's "Modified Gauss-Newton Method" to allow for constraints on θ.

(5) The removal of some or all of the approximations. (This is really the original problem, but is extremely more difficult since, for one thing, it precludes the use of a compact matrix notation).

(6) A purely geometric approach; only for a special class of problems. (The analogy is the graphical solution of a set of simultaneous equations).

This report indicates that progress has been made toward the solution of a restricted class of constrained estimation problems. The avenue of approach used here has proved promising and will be pursued. Some specific problems have been solved(see also the section of this report where the CER for CIT was derived) and results applicable to the cost model problem are expected. Specifically, the area(3) above can be applied to the fitting of a cubic equation subject to the restriction that it be a monotonic increasing function (representing a cumulative distribution function - used in the subjective probability problem being investigated). In fact, a theoretically sound solution has just been obtained for that cubic problem. This theory is now being used to construct a usable algorithm suitable for incorporation into the Hartley-Hocking algorithm.

REFERENCES

[1] Strand, T.G., Kohl, D.A., and Bonham, R.A., "Modification to the Newton-Raphson Method for the Fitting of Nonlinear Functions by Least Squares", The Journal of Chemical Physics, Vol. 39, No. 5, September, 1963, pp. 1307-1310.

[2] Marquardt, Donald W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", Journal of the Society for Industrial and Applied Mathematics, Vol. 11, 1961, pp. 431-441.

[3] Hartley, H.O., "Modified Gauss-Newton Method for the Fitting of Non-linear Regression Functions by Least Squares", Technometrics, Vol. 3, 1961, pp. 269-80.

[4] Hartley, H.O., and Hocking, R.R., "Convex Programming by Tangential Approximation", Management Science, Vol. 9, No. 4, July, 1963, pp. 600-612.

PART II

DEVELOPMENT OF COST ESTIMATING RELATIONSHIPS TO BE USED IN THE SPACECRAFT COST MODEL

By

William P. Cooke

For

Industrial Engineering Department

Texas A&M University

May 1966

INTRODUCTION

During the review of Cost Estimating Relationship (CER) work as part of the cost research grant the recurring problem was that of interaction among the factors and the resultant effect upon the reproduction of the costs used in developing the original CER's. It was desirable to use certain functional forms which were derived through the use of expertise to represent first order effects. The higher order interactions which were used were convenient and are not the only ones that could be used.

The important point is the methodology developed. The cost data, the functional forms, the variables, and the number of parameters and higher order terms are minor points compared to the concept and the methodology evolved in this research.

The reference to specific organizations is not critical or significant since the development is independent of the work performed by those organizations. Their work was simply used as a data base in order that the discussion could be phrased in familiar terms.

ASSUMPTIONS

The procedures and results presented in this report are based on the following assumptions:

1. The lack of fit to the observed data of the CER's developed by Booz-Allen is due to:

   (a) No consideration of "interaction" among the predictor variables.

   (b) Subjective weighting of the single-variable "predictors".

2. The four variables selected as predictor variables are valid; i.e., that Booz-Allen and NASA are in agreement that these four variables can be used to predict subsystem costs satisfactorily.

3. The basic premise of Booz-Allen; that the functional form of the CER's involves natural logarithms of the variables, is preserved.

4. There is some function of the four variables which represents subsystem cost, and which can be adequately approximated by a polynomial function in the logarithms of the variables.

A COST ESTIMATING RELATIONSHIP FOR CIT

The procedure for constructing a CER with a minimum of data will be presented in the form of an example, i.e., the actual construction of the CER for CIT for subsystems (Structure-Primary).

The basic tool for estimating parameters in the prescribed functions will be be least-squares, and in fact may be termed "constrained successive least-squares" for reasons which will become apparent. The minimum of data available precludes estimating variances of the parameter estimates, so the criterion of validity of the CER will be the "predictability" of the observed costs by the CER.

The procedure follows:

(1) Single-Variable Predictors

We make the following identifications:

X_1 = SW = Subsystem weight

X_2 = MG = Maximum mission g's

X_3 = ML = Module length

X_4 = NF = Number of flight functions.

Suppose that CIT can be predicted by each individual variable in the form

$$CIT_i = K_i \ln (X_i + 1), \ i = 1, 2, 3, 4.$$

Applying least-squares calculations to the data, we obtain

$$CIT_1 = 583 \ln (X_1 + 1)$$

$$CIT_2 = 1{,}768 \ln (X_2 + 1)$$

$$CIT_3 = 1{,}592 \ln (X_3 + 1)$$

$$CIT_4 = 2{,}289 \ln (X_4 + 1)$$

Keeping in mind that we are trying to estimate, in thousands of dollars, CIT, where

$$CIT = \begin{bmatrix} 5{,}016 \\ 4{,}730 \\ 4{,}817 \end{bmatrix} = \begin{matrix} \text{CIT for Gemini} \\ \text{CIT for Apollo LEM} \\ \text{CIT for Apollo C\&S} \end{matrix} ,$$

we write out the single-variable "predictors" as

$$\hat{CIT}_1 = \begin{matrix} 4{,}502 \\ 4{,}612 \\ 5{,}357 \end{matrix} , \quad \hat{CIT}_2 = \begin{bmatrix} 5{,}060 \\ 3{,}885 \\ 5{,}383 \end{bmatrix} , \quad \hat{CIT}_3 = \begin{bmatrix} 4{,}813 \\ 4{,}422 \\ 5{,}259 \end{bmatrix} , \quad \hat{CIT}_4 = \begin{bmatrix} 4{,}760 \\ 4{,}760 \\ 5{,}029 \end{bmatrix}$$

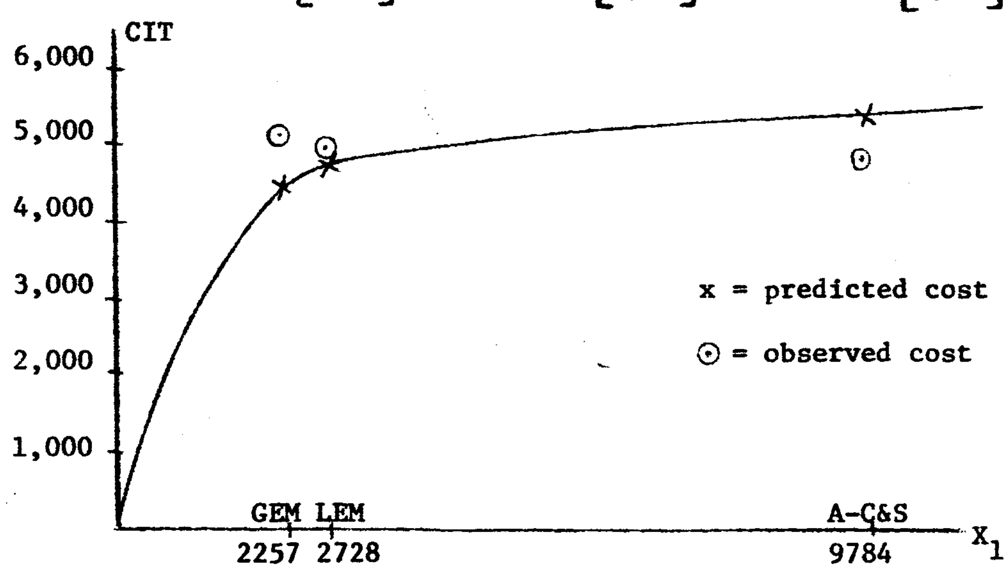


Figure 1. CIT_1: Estimated by Weight Only.

The figure above indicates the form of the prediction equations as well as their lack of fit to the observed data.

Note that a simple subjective weighting of these 4 "predictors" will produce an equation similar to that derived by Booz-Allen. Note also that there is no weighting scheme that would result in an adequate prediction of CIT for Apollo C&S. The closest we could come to the 4,817 would be 5,029.

(2) Two-Variable Predictors

We now propose to form all possible predictors based on pairs of the four variables [there will be $\binom{4}{2} = 6$ such equations] by first forming weighted sums of the single-variable predictors and then adjusting by interaction-type terms of a quadratic form. The resulting equations will be second degree polynomials in the two variables, which begins the construction of the polynomial suggested in assumption (4).

A sample calculation, using variables X_1 and X_2, follows.

Form $CIT = p\,\hat{CIT}_1 + (1-p)\,\hat{CIT}_2$, $o \le p \le 1$ and obtain the constrained least-squares estimate of p (restricted by $0 \le p \le 1$). This is an attempt to predict CIT exactly by a weighted sum of CIT_1 and CIT_2. We discover, however, that it does not do so, but call this a new estimator CIT^*_{12}. Our caluulation yields p = .78, so we have

$$CIT^*_{12} = .78\,CIT_1 + .22\,CIT_2\,, \quad \text{and}$$

$$\hat{CIT}^*_{12} = \begin{bmatrix} 4,625 \\ 4,452 \\ 5,363 \end{bmatrix} .$$

Now there is an obvious discrepancy in the actual and predicted costs, which we try to account for by interaction terms, hypothesizing that

$$CIT = CIT_{12} = .78\,CIT_1 + .22\,CIT_2 + a\ln^2(X_1+1)$$
$$+ b\ln^2(X_2+1) + c\ln(X_1+1)\ln(X_2+1).$$

A least-squares solution for the parameters a, b, and c yields the two-variable predictor

$$CIT_{12} = CIT^*_{12} + 590 \ln^2 (X_1 + 1) + 5,754 \ln^2 (X_2 + 1)$$
$$- 3,708 \ln (X_1 + 1) \ln (X_2 + 1).$$

Similar calculations yield the other five two-variable predictors which, along with the one above, are shown below.

$$CIT^*_{12} = .78\ CIT_1 + .22\ CIT_2$$
$$CIT^*_{13} = CIT_3$$
$$CIT^*_{14} = CIT_4$$
$$CIT^*_{23} = CIT_3$$
$$CIT^*_{24} = .03\ CIT_2 + .97\ CIT_4$$
$$CIT^*_{34} = CIT_4$$

After adjusting for interactions, we obtain the final two-variable predictors:

$$CIT_{12} = CIT^*_{12} + 590 \ln^2 (X_1 + 1) + 5,754 \ln^2 (X_2 + 1)$$
$$- 3,708 \ln (X_1 + 1) \ln (X_2 + 1).$$
$$CIT_{13} = CIT^*_{13} + 5,312 \ln^2 (X_1 + 1) + 38,496 \ln^2 (X_3 + 1)$$
$$- 28,630 \ln (X_1 + 1) \ln (X_3 + 1).$$
$$CIT_{14} = CIT^*_{14} + 1,407 \ln^2 (X_1 + 1) + 22,734 \ln^2 (X_4 + 1)$$
$$-11,330 \ln (X_1 + 1) \ln (X_4 + 1).$$
$$CIT_{23} = CIT^*_{23} + 19,820 \ln^2 (X_2 + 1) + 14,975 \ln^2 (X_3 + 1)$$
$$- 34,559 \ln (X_2 + 1) \ln (X_3 + 1).$$
$$CIT_{24} = CIT^*_{24} - 38,798 \ln^2 (X_2 + 1) - 56,646 \ln^2 (X_4 + 1)$$
$$+ 94,605 \ln (X_2 + 1) \ln (X_4 + 1).$$
$$CIT_{34} = CIT^*_{34} - 14,534 \ln^2 (X_3 + 1) - 28,912 \ln^2 (X_4 + 1)$$
$$+ 41,058 \ln (X_3 + 1) \ln (X_4 + 1).$$

The "predictions" of CIT given by these are:

$$\hat{CIT}_{12} = \begin{bmatrix} 4,990 \\ 4,703 \\ 4,781 \end{bmatrix}, \quad \hat{CIT}_{13} = \begin{bmatrix} 5,037 \\ 4,767 \\ 4,816 \end{bmatrix}, \quad \hat{CIT}_{14} = \begin{bmatrix} 5,031 \\ 4,735 \\ 4,832 \end{bmatrix},$$

$$\hat{CIT}_{23} = \begin{bmatrix} 5,010 \\ 4,728 \\ 4,818 \end{bmatrix}, \quad \hat{CIT}_{24} = \begin{bmatrix} 5,054 \\ 4,735 \\ 4,800 \end{bmatrix}, \quad \hat{CIT}_{34} = \begin{bmatrix} 5,019 \\ 4,757 \\ 4,859 \end{bmatrix},$$

which are each trying to predict $CIT = \begin{bmatrix} 5,016 \\ 4,730 \\ 4,817 \end{bmatrix}$.

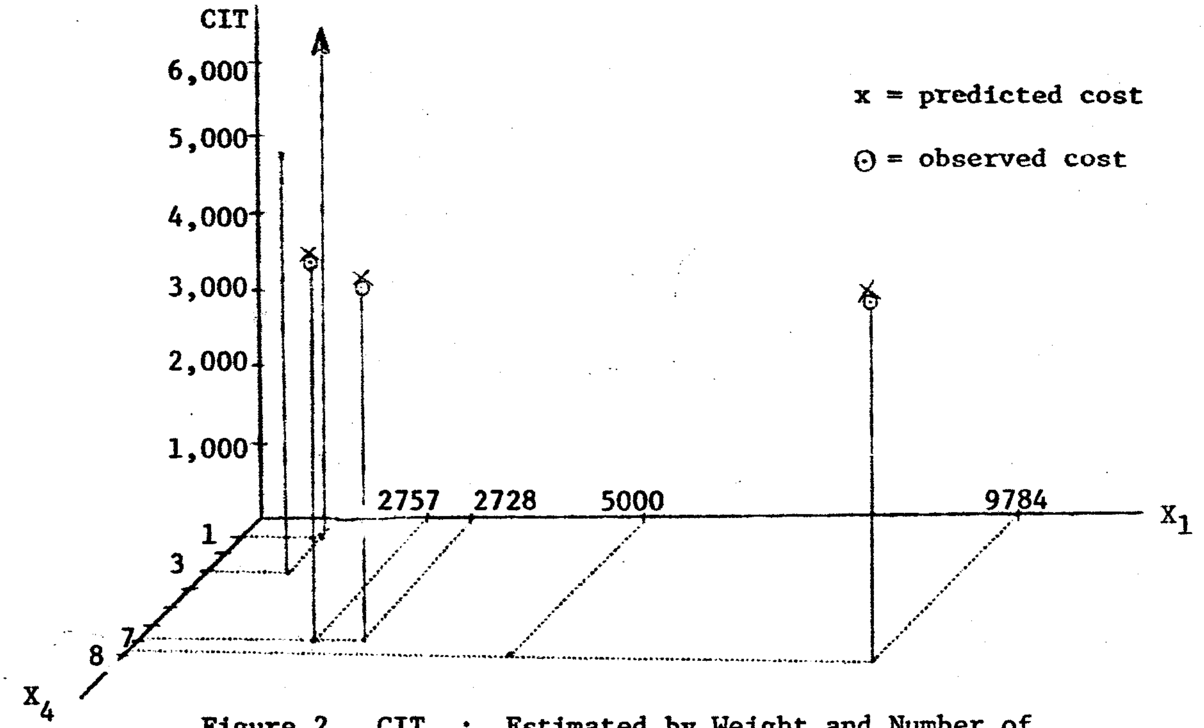


Figure 2. CIT_{14}: Estimated by Weight and Number of Flight Functions

Any one of these is a fairly good predictor of CIT. The figure illustrates the nature of the surface on which predicted costs, as a function of only two variables, will lie. However, we would like an equation which tries to account for the contribution of all four variables, since it has been decieed [assumption (2)] that each is important as a predictor.

(3) Three-Variable Predictors

Continuing the same approach, we form weighted sums of the two-variable predictors. Here the weighting restrictions pose interesting but simple non-linear programming problems, since we have problems of the type: Minimize

$$[CIT - p_1\, CIT_{12} - p_2\, CIT_{13} - (1 - p_1 - p_2)\, CIT_{23}]^2$$

subject to the restrictions $o \leq p_1 \leq 1$, $o \leq p_2 \leq 1$, $o \leq p_1 + p_2 \leq 1$. Fortunately the problem has a simple geometric solution which can be converted to a set of logical instructions that can be handled by a computer without recourse to convex programming algorithms. These sample calculations were done on a desk calculator to get a "feel" for the problem; but future CER's are expected to come from a single computer program incorporating the above procedures.

The results of the calculations are:

$$CIT^*_{123} = .11\ CIT_{13} + .89\ CIT_{23}$$

$$CIT^*_{124} = .91\ CIT_{14} + .09\ CIT_{24}$$

$$CIT^*_{134} = CIT_{14}$$

$$CIT^*_{234} = .84\ CIT_{23} + .13\ CIT_{24} + .03\ CIT_{34}\ .$$

At this stage an element of subjectivity enters into the problem; that of adjusting for three-factor interactions. The data are not sufficient to allow all possible three-factor interactions to be used, so it was decided to use only $(\ln)^3$ terms. The final three-variable predictors are then:

$$CIT_{123} = CIT^*_{123} - .0274 \ln^3 (X_1 + 1) + .0674 \ln^3 (X_2 + 1) + .5079 \ln^3 (X_3 + 1).$$

$$CIT_{124} = CIT^*_{124} + .0281 \ln^3 (X_1 + 1) - .8513 \ln^3 (X_2 + 1) - 1.0134 \ln^3 (X_4 + 1).$$

$$CIT_{134} = CIT^*_{134} + .0193 \ln^3 (X_1 + 1) - 2.1066 \ln^3 (X_3 + 1) + 3.9616 \ln^3 (X_4 + 1).$$

$$CIT_{234} = CIT^*_{234}$$

These yield the "predictions":

$$\hat{CIT}_{123} = \begin{bmatrix} 5,016 \\ 4,730 \\ 4,817 \end{bmatrix},\ \hat{CIT}_{124} = \begin{bmatrix} 5,017 \\ 4,731 \\ 4,816 \end{bmatrix},\ \hat{CIT}_{134} = \begin{bmatrix} 5,017 \\ 4,735 \\ 4,813 \end{bmatrix},\ \hat{CIT}_{234} = \begin{bmatrix} 5,016 \\ 4,730 \\ 4,817 \end{bmatrix},$$

which hardly deviate at all from the observations.

(4) Four-Variable Predictor

At this stage it was observed that a correct weighting of these predictors must yield $CIT_{1234} = CIT_{234}$, which would ignore X_1 completely. This could lead to an intriguing conjecture (ignore weight?), but would not seem to make sense practically. It would seem that the more logical approach, to carry our line of reasoning to completion, would be to weight the hardly distinguishable three-variable predictors equally, and make a final four-factor interaction adjustment.

This yields

$$CIT^*_{1234} = .25\ [CIT_{123} + CIT_{124} + CIT_{134} + CIT_{234}] \quad ,$$

and finally

$$CIT_{1234} = CIT^*_{1234} - .0024 \ln (X_1 + 1) \ln (X_2 + 1) \ln (X_3 + 1) \ln (X_4 + 1) \quad ,$$

where

$$\hat{CIT}_{1234} = \begin{bmatrix} 5,016 \\ 4,731 \\ 4,816 \end{bmatrix} \quad .$$

(5) <u>The CER for CIT</u>

Performing the arithmetic of combining all the previous results exhibits the CER in terms of the original variables.

$$\begin{aligned} CIT_{1234} = \{ & 2.9172 \ln (X_2 + 1) + 732.32 \ln (X_3 + 1) + 1,232.3 \ln (X_4 + 1) \\ & + 817.92 \ln^2(X_1 + 1) + 6,438.26 \ln^2(X_2 + 1) \\ & + 7,426.32 \ln^2(X_3 + 1) + 7,523.11 \ln^2(X_4 + 1) \\ & - 787.33 \ln (X_1 + 1) \ln (X_3 + 1) \\ & - 5,410.08 \ln (X_1 + 1) \ln (X_4 + 1) \\ & - 14,946.77 \ln (X_2 + 1) \ln (X_3 + 1) \\ & + 5,203.28 \ln (X_2 + 1) \ln (X_4 + 1) \\ & + 307.94 \ln (X_3 + 1) \ln (X_4 + 1) \} \\ & + .0050 \ln^3(X_1 + 1) - .1960 \ln^3(X_2 + 1) - .3997 \ln^3(X_3 + 1) \\ & + .7371 \ln^3(X_4 + 1) \\ & - .0024 \ln (X_1 + 1) \ln (X_2 + 1) \ln (X_3 + 1) \ln (X_4 + 1) \ . \end{aligned}$$

(6) <u>Discussion of CER and Recommendations</u>

Observe that the part of the CER set off in braces ({}) is a quadratic form in the four variables, and represents an approximation of only second degree to the theoretical (hypothesized) function which actually yields CIT for any combination of the four variables. Ignoring the higher degree terms completely yields the prediction

$$\begin{bmatrix} 5{,}037 \\ 4{,}732 \\ 4{,}824 \end{bmatrix},$$

that is, the higher degree terms do not make a large contribution <u>for this range of the data</u>. If the expected range of future input variable X_1, X_2, X_3, X_4 is to be near the range used in deriving the CER, then it may be expected that the quadratic form only would be a fairly good "predictor".

Such an observation allows us to make the statement that the predictor is "nice" in the respect that it leads us to no obviously contradictory cost predictions such as negative cost, maximum cost, etc.

If the quadratic form (ignoring terms of degree 3 and higher in the CER) is investigated for maxima and minima, the result is that it has a minimum where

$$\begin{array}{lcl} X_1 + 1 \doteq \exp[-293] < 1 & & X_1 < o \\ X_2 + 1 \doteq \exp[-84] < 1 & \text{implying} & X_2 < o \\ X_3 + 1 \doteq \exp[-98] < 1 & & X_3 < o \\ X_4 + 1 \doteq \exp[-74] < 1 & & X_4 < o \end{array},$$

which is outside the range of feasible data.

The conclusion is that the <u>restricted</u> CER has no maximum within the feasible range (positive X's) of the data, and the possibility of a "negative cost" appears to exist only in the neighborhood of $X_1 = o$, $X_2 = o$, $X_3 = o$, $X_4 = o$, a point of no interest (except that cost in zero there).

The <u>full</u> (all terms) CER has been investigated throughout the anticipated range of input variables by simply programming the equation and generating cost predictions for various combination of the variables.

A discussion of the program and reasons for its use will be found in another section of this report. The figure below represents graphically the results of that investigation. It is a three-dimensioned (holding X_4 fixed at $X_4 = 8$) contour representation of costs, where the costs are indicated within the flags drawn at various points representing inputs (X_1, X_2, X_3).

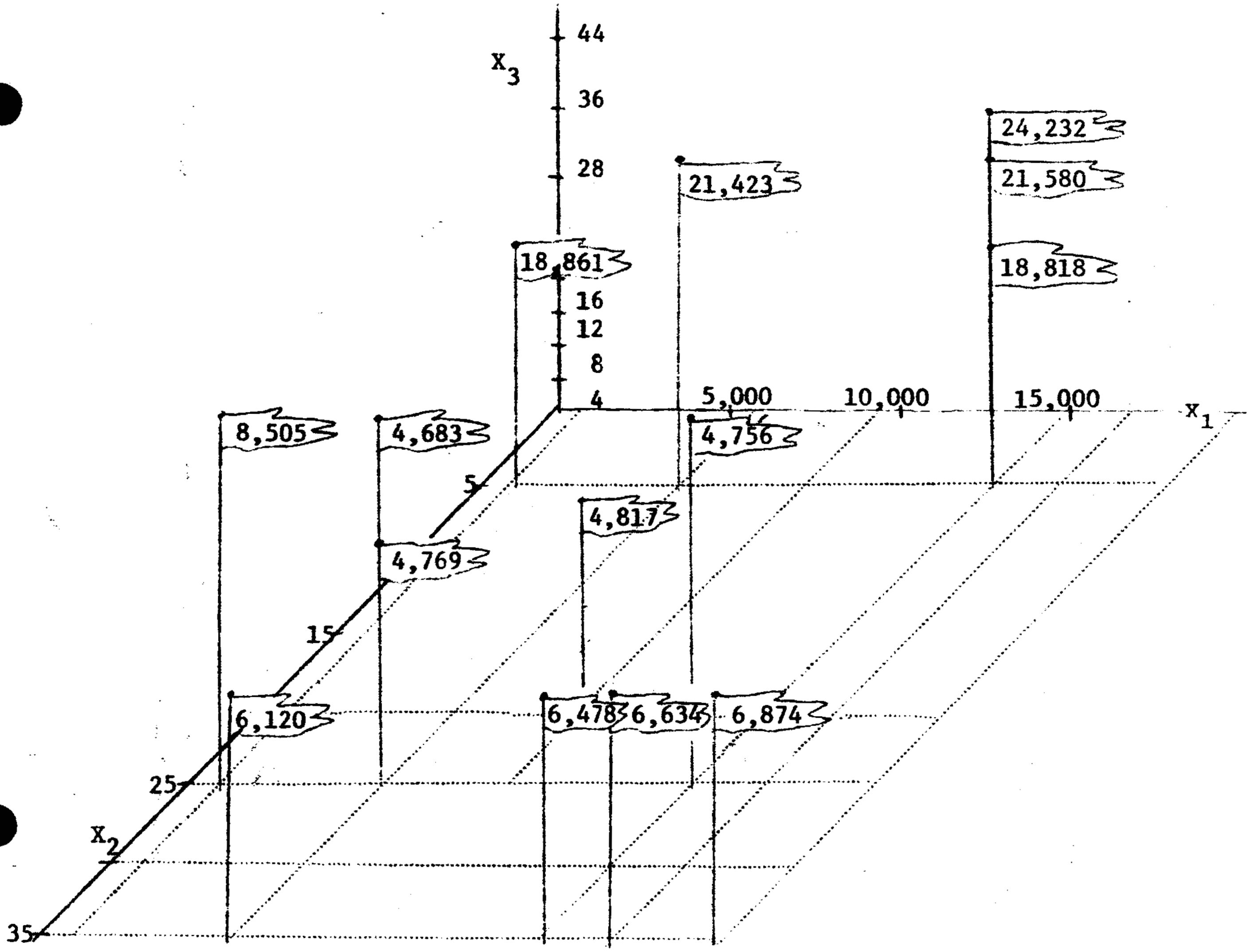


Figure 3. CIT_{1234}: Complete CER, X_4 Fixed at X_4 = 8.

While the figure indicates that no results contrary to common sense were obtained, the recommendation is that the quadratic form in the CER appears to be an adequate predictor for CIT for points not too discrepant from the range of data used in the development of the CER. This would essentially maintain Booz-Allen's original four variables and logarithmic form of the function, while increasing the degree of the polynomial by one.

(7) Comments on the Number of Terms in the CER

The CER is a fourth degree polynomial form in four variables and has 17 terms, arising from the nature of the method used (constrained least-squares) and from the minimum of data that was available for constructing the function.

An analogy can be made with the standard statistical procedure of finding the degree of a polynomial (in one variable) that adequately describes a set of data when the actual functional form is unknown. In that procedure, a linear function is first fitted, tested for significance of the coefficient of the first degree term, and, if significant, a polynomial of second degree is fitted, tested, and so on until no significance is obtained for the coefficient of the term of highest degree.

The method proposed in this report utilizes no significance tests because of the lack of available data. Our criterion for deciding upon the degree of the polynomial to be used is simply to continue until the data have been fit essentially exactly and then say that, since the polynomial describes the data perfectly, we have an adequate fit. However, because of the few data points, after reaching the stage of a second degree polynomial a decision had to be made as to what <u>kind</u> of third degree terms would be used, these being 20 different ones to choose from. Likewise for the fourth degree terms, where there are 29 choices. These decisions were necessarily subjective and were based only on using terms which improved the fit. Hence the recommendation that the quadiatic form might be an adequate predictor (it required no such decisions).

It should be noted that if it were decided that a fourth degree polynomial in the 4 variables is desired, and that we should fit a <u>general</u> (all possible terms) fourth degree polynomial, then 63 data points would be required if standard least-squares is to be used, since such a polynomial will have 63 different terms. Even if that much data is available, the estimation task is formidable, requiring the solution of a system of 63 simultaneous equations in 63 unknowns! In such a situation it is usually decided to attempt to describe the data with a quadratic form in the variables. This general form would contain 14 terms, however, so standard least-squares cannot be used even here, since we do not have that much data.

USE OF THE METHOD FOR OTHER CER'S

It appears reasonable to say that the method used in constructing CIT_{1234} would work for the other desired CER's, given that the original assumptions are acceptable to NASA.

As an example, a sample two-variable predictor for CDE was calculated, and yielded

$$\begin{aligned} CDE_{12} = {} & 11{,}019 \ln (X_1 + 1) - 53{,}516 \ln^2 (X_1 + 1) \\ & - 550{,}805 \ln^2 (X_2 + 1) \\ & + 345{,}786 \ln (X_1 + 1) \ln (X_2 + 1), \end{aligned}$$

so that

$$\hat{CDE}_{12} = \begin{bmatrix} 24{,}216 \\ 89{,}235 \\ 150{,}719 \end{bmatrix},$$

while the observation actually made was

$$CDE = \begin{bmatrix} 24{,}185 \\ 89{,}205 \\ 150{,}681 \end{bmatrix},$$

leading us to expect the same "goodness of fit" for a CER for CDE based on the same method.

The procedure will be programmed for computer and, especially if the quadratic form only is desired, should be a fairly simple program to write (the logic for the non-linear programming aspect would be quite simple). As soon as this is done, all the CER's could be constructed, investigated for "reasonableness", and submitted to NASA.

It is felt worthwhile to re-emphasize that the use of only the quadratic form (a) should lead to CER's with good properties and (b) removes the element of subjectivity (apart from the original assumptions) because no decisions as to <u>type</u> of three and four factor interaction need be made.

USE OF OTHER FUNCTIONAL FORMS: CONCLUSION

It is of course apparent that the method depends on the assumptions made at the beginning of this section of the report. It must be pointed out that, given a "free hand" as to the form of the functions used, we may arrive at functions which predict the observed points closely, and could do so with only one variable.

For example, we could form

$$CDE_3 = 67{,}403\ X_3 - 6{,}416\ X_3^2 + 155\ X_3^3 \quad ,$$

yielding

$$\hat{CDE}_3 = \begin{bmatrix} 23{,}652 \\ 88{,}915 \\ 149{,}393 \end{bmatrix} \quad .$$

This gives a fairly good fit, and is simply a cubic equation forced through $X_3 = 0$ and involving only the single variable X_3 .

It is obvious that an infinity of such functions might be hypothesized for any one CER; each doing a good job of "prediction".

The objections to this approach are (a) the problem would have as many "solutions" as there were people to try to solve it, and (b) we would be basing the form of the function on the configuration of the data rather than on past experience and consideration of the nature of the costs themselves; a dangerous practice at best.

By restricting ourselves to a set of assumptions that can be at least partially justified by experience, we have a rational foundation on which to build our model.

PART III

COMPUTERIZED EVALUATION OF COST ESTIMATING RELATIONSHIPS

By

Jack C. Britian

For

Industrial Engineering Department

Texas A&M University

May 1966

COMPUTERIZED EVALUATION OF COST ESTIMATING RELATIONSHIPS

The purpose of the Computer Program is to evaluated the Cost Estimating Relationships of the form discussed below. The specific CER that will be discussed is for the initial tooling cost. In this case the variable in the prediction equation are x_1, subsystem weight; x_2, maximum g's; x_3, module length; and x_4, the number of flight functions.

The predictor or CER developed for the initial tooling cost is:

$$\begin{aligned} ITC\ (x_1, x_2, x_3, x_4) = {} & 2.9172 \ln (x + 1) + 732.32 \ln (x_3 + 1) + 1,232.3 \ln(x_4 + 1) \\ & + 817.92 \ln^2 (x_1 + 1) + 6,438.26 \ln^2 (x_2 + 1) \\ & + 7,426.32 \ln^2 (x_3 + 1) + 7,523.11 \ln^2 (x_4 + 1) \\ & - 7,87.33 \ln (x_1 + 1) \ln (x_3 + 1) - 5,410.08 \ln (x_1 + 1) \cdot \\ & \cdot \ln (x_4 + 1) - 14,946.77 \ln (x_2 + 1) \ln (x_3 + 1) \\ & + 5,203.28 \ln (x_2 + 1) \ln (x_4 + 1) + 307.94 \ln (x_3 + 1) \cdot \\ & \cdot \ln (x_4 + 1) + .0050 \ln^3 (x_1 + 1) - 0.1960 \ln^3 (x_2 + 1) \\ & - .3997 \ln^3 (x_3 + 1) + .7371 \ln^3 (x_4 + 1) \\ & - .0024 \ln (x_1 + 1) \ln (x_2 + 1) \ln (x_3 + 1) \ln (x_4 + 1) \end{aligned}$$

Consideratiin was given to evaluating the ITC predictor for stationary points analytically. This approach to the problem leads to differentiating with respect to $\ln (x_i + 1)$, $i = 1,2,3,4$ and setting each of the four equations equal to zero, and solving for the desired variables (x_1, x_2, x_3, and x_4).

The system of equation resulting from taking the partial derivative and setting them equal to zero becomes:

$$\begin{aligned} \frac{\partial(ITC)}{\partial[\ln(x_1 + 1)]} = {} & 1635.84 \ln (x_1 + 1) - 787.33 \ln (x_3 + 1) - 5,410.08 \ln (x_4 + 1) \\ & + .015 \ln^2 (x_1 + 1) - 0.0024 \ln (x_2 + 1) \ln (x_3 + 1) \ln (x_4 + 1) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial(ITC)}{\partial[\ln(x_2 + 1)]} = {} & 2.9172 + 12,876.52 \ln (x_2 + 1) - 14,946.77 \ln (x_3 + 1) \\ & + 5,203.28 \ln (x_4 + 1) - .5880 \ln^2 (x_2 + 1) - 0.0024 \ln (x_1 + 1) \cdot \\ & \cdot \ln (x_3 + 1) \ln (x_4 + 1) = 0 \end{aligned}$$

$$\frac{\partial(\text{ITC})}{\partial[\ln(x_3+1)]} = 732.32 + 14{,}852.64 \ln(x_3+1) - 787.33 \ln(x_1+1) - 14{,}946.77 \ln(x_2+1) + 307.94 \ln(x_4+1) - 1.1991 \ln^2(x_3+1) - 0.0024 \ln(x_1+1)\ln(x_2+1)\ln(x_4+1) = 0$$

$$\frac{\partial(\text{ITC})}{\partial[\ln(x_4+1)]} = 1{,}232.3 + 15{,}046.22 \ln(x_4+1) - 5{,}410.08 \ln(x_1+1) + 5{,}203.28 \ln(x_2+1) + 307.94 \ln(x_3+1) + 2.2113 \ln^2(x_4+1) - 0.0024 \ln(x_1+1)\ln(x_2+1)\ln(x_3+1) = 0$$

An analytical solution to this system of four equations for stationary (maximum, minimum, or saddle) points was not readily apparent. For this reason an iterative solution to the system of equations was necessary.

A FORTRAN program was written to examine the CER, taking advantage of the computers speed and accuracy, for any unreasonable behavior. This technique of examining the CER allowed many more data points to be examined than otherwise would have been possible.

The FORTRAN program was designed such that all coefficients and data were loaded on data cards (Standard IBM 5081, 80 column cards). See Figure 2. The data cards, supplying the values of the coefficients used in the CER, utilize all eighty columns of two cards and ten columns of a third card. The second type of data card used utilizes the first forty columns to supply the values of x_i, $i = 1,2,3,4$ (subsystem weight, etc.). Columns sixty-one (61) thru and including seventy-two (72) may be used for comment or marking of the data as to the type of spacecraft (Gemini, Apollo, etc.).

The program was designed for maximum utilization of the memory of the IBM 7094 so that the program, if necessary, can be expanded to make all calculations necessary in the development of the CER.

The output of the FORTRAN program lists, first, the values of the coefficients used in the CER. Secondarily, the program lists the data (x_1, x_2, x_3, x_4) used for each iteration adjacent to the value for the ITC obtained using the data.

The listing of the data used for each value of the CER facilitates the investigation by helping to eliminate error in plotting the points.

It can be seen from the flow chart of the program that other advantages were obtained by the use and design of the computer program. One feature that proves useful is that higher order terms of the predictor can easily be deleted if the quadratic form is an adaquate predictor of costs.

The ability of deletion without changing the entire program contributes to the fact that the program can easily be used to compare the predictor based on lower degree terms to the predictor based on the complete predictor equation.

The data used to evaluate the CER was such that the predictor could be examined over all possible combinations of available and theoritical data. The theoritical data was designed such that it progressed in logical and convenient increments from the lowest available data values to approximately two or three times the highest available data values.

The CER was then calculated for each of these data sets (x_1, x_2, x_3, x_4) and hand plotted in a three dimensional contour coordinate system for observation. The section on "Development of a Cost Estimating Relationship..." contains an illustration of these results.

This type of program should be used as a checkout procedure on each CER and/or combination of CER's. Checking each mathematical model in this manner would call attention to the reliability of the CER and the independence of the CER's when grouped within a model.

The input/output and program are an addendum to this section of the report.

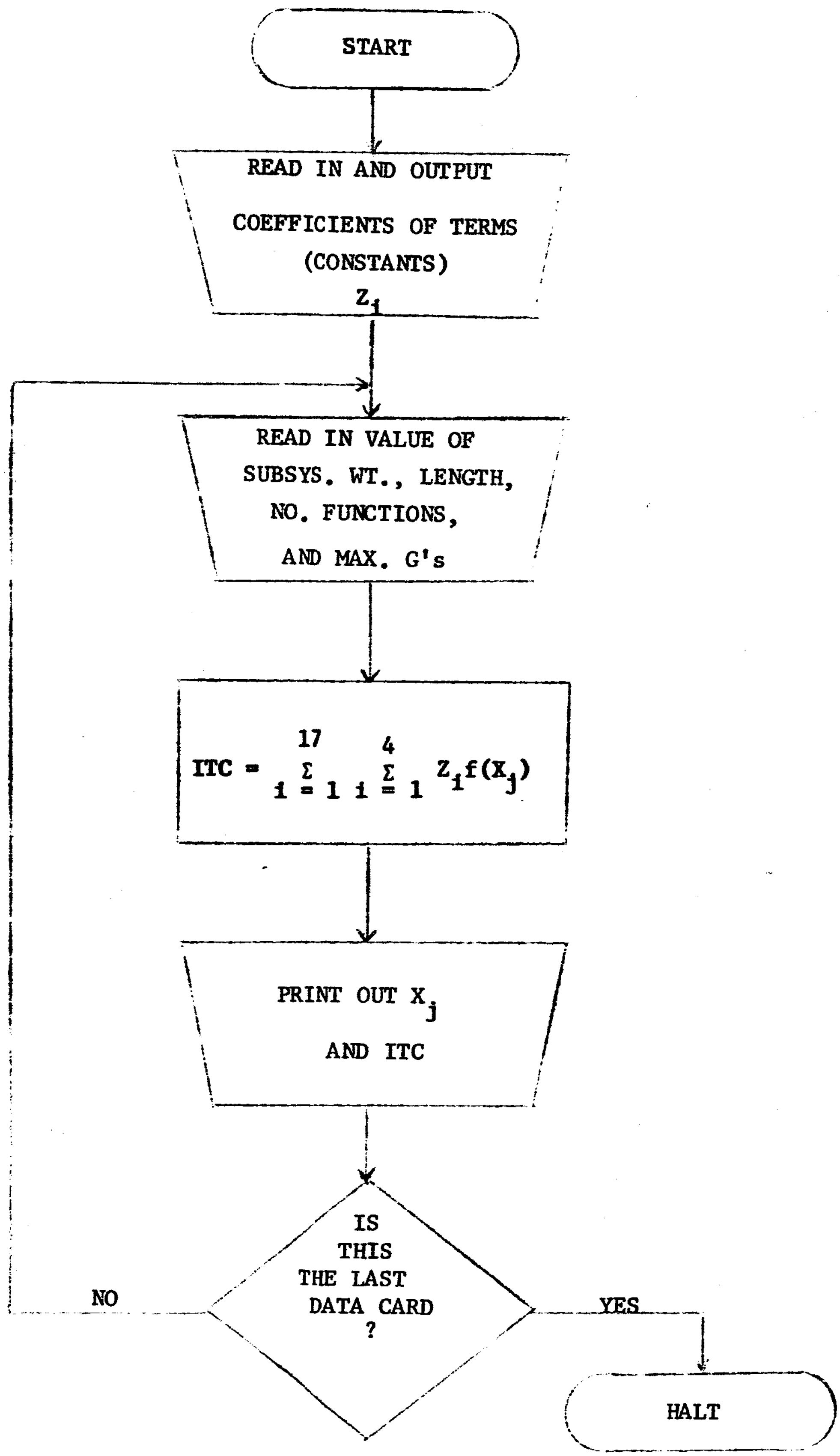


FIGURE 1. FLOW CHART OF CER EVALUATION PROGRAM

INPUT FORMAT COLUMN IDENTIFICATION

	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	
COEFFICIENTS OF TERMS IN CER	Z(1)	Z(2)	Z(3)	Z(4)	Z(5)	Z(6)	Z(7)	Z(8)	CARD 1
	Z(9)	Z(10)	Z(11)	Z(12)	Z(13)	Z(14)	Z(15)	Z(16)	CARD 2
	Z(17)	-	-	-	-	-	-	-	CARD 3

	1–10	11–20	21–30	31–40	61–72	
DATA	X(1)	X(2)	X(3)	X(4)	COMMENT	N CARDS

EXAMPLE OF INPUT DATA

2.917200	732.3200	1232.300	817.9200	6438.260	7426.320	7523.110	787.3300
5410.000	14946.77	5203.280	307.940	0.005000	0.196000	0.399700	0.7371000
0.002400							

2728.00	8.00	15.08	7.00			GEMINI
2257.00	7.00	19.56	16.50			L.E.M.

EXAMPLE OF PROGRAM LISTING

```
      DIMENSION X(4),Z(17),NAME(2)
      REAL ITC
      READ(5,2)(Z(I),I = 1,17)
    2 FORMAT( 8F10.5)
      WRITE(6,15)
   15 FORMAT(1H1,56X,20H COST MODEL RESULTS ///56X,23HX(1) = SUBSYSTEM W
     1EIGHT,/56X,18HX(2) = MAXIMUM G-S,/,56X,20HX(3) = MODULE LENGTH,/,5
     16X,26HX(4) = NUMBER OF FUNCTIONS,)
      WRITE(6,10)(I ,  Z(I),I = 1,17)
   10 FORMAT(1H0,10X,3H Z(,I2,5H ) = ,F10.4)
   99 READ(5,1)(X(I),I = 1,4),(NAME(I),I = 1,2)
    1 FORMAT(4F10.2,20X,2A6)
      A = ALOG(X(2) + 1.0)
      B = ALOG(X(3) + 1.0)
      C = ALOG(X(4) + 1.0)
      D = ALOG(X(1) + 1.0)
      D2 = D**2
      E = A**2
      F = B**2
      G = C**2
      H = D*B
      O = D*C
      P = A*B
      Q = A*C
      R = B*C
      S = D**3
      T = A**3
      U = B**3
      V = C**3
      W = A*D*B*C
      ITC = Z(1)*A + Z(2)*B + Z(3)*C + Z(4)*D2 + Z(5)*E + Z(6)*F + Z(7)*
     1G - Z(8)*H - Z(9)*O - Z(10)*P + Z(11)*Q + Z(12)*R + Z(13)*S - Z(14
     2)*T - Z(15)*U + Z(16)*V - Z(17)*W
      WRITE (6,11)( I,  X(I),I = 1,4)
   11 FORMAT(//10X, 3H X(,I2,5H ) = ,F10.2)
      WRITE(6,12)ITC,(NAME(I),I = 1,2)
   12 FORMAT(//,10X,36HTHE INITIAL TOOLING COST(1,2,3,4) = ,F15.2,10X,2A
     16)
      GO TO 99
      END
$DATA
```

EXAMPLE OF OUTPUT LISTING

```
                         COST MODEL RESULTS

                         X(1) = SUBSYSTEM WEIGHT
                         X(2) = MAXIMUM G-S
                         X(3) = MODULE LENGTH
                         X(4) = NUMBER OF FUNCTIONS

Z( 1) =     2.91720
Z( 2) =   732.32000
Z( 3) =  1232.29999
Z( 4) =   817.92000
Z( 5) =  6438.25995
Z( 6) =  7426.31995
Z( 7) =  7523.10999
Z( 8) =   787.32999
Z( 9) =  5410.00000
Z(10) =14946.76990
Z(11) =  5203.27997
Z(12) =   307.94000
Z(13) =     0.00500
Z(14) =     0.19600
Z(15) =     0.39970
Z(16) =     0.73710
Z(17) =     0.00240

X(1) =        2728.00
X(2) =           8.00
X(3) =          15.08
X(4) =           7.00
THE INITIAL TOOLING COST(1,2,3,4) =            5016.77    GEMINI

X(1) =        2257.00
X(4) =           7.00
X(3) =          19.56
X(2) =          16.50
THE INITIAL TOOLING COST(1,2,3,4) =            4731.70    L.E.M.
```

PART IV

DYNAMIC PROGRAMMING ALGORITHM FOR DETERMINING 'BEST FIT'

By

Glen D. Self

For

Industrial Engineering Department

Texas A&M University

May 1966

DYNAMIC PROGRAMMING ALGORITHM FOR DETERMINING 'BEST FIT'

The problem of weighting individual predictors of total subsystem cost in order to combine predictors of the same value into a single predictor is one which can be approached by more than one method. One method which was reviewed as part of the services performed by Texas A&M was that of subjectively weighting the individual predictors which contained a single independent variable according to the importance that variable was felt to have on the total cost function being considered. This particular approach did not produce consistently good results and did not present a quantitative base for making decisions as to whether the individual predictors were at fault or the weighting scheme being used. Therefore, by expanding the problem to (1) one of having any number of terms that were to be combined into a single function under the constraint that the sum of the weights used should approximate unity (2) one being further restricted by a small number of data points (which precipitated the original problem of only using a single variable predictor for a given type of function and still have some degrees of freedom associated with the error sum of squares) (3) one having an inherent flexibility such that consecutive last terms could be deleted from consideration and still provide the optimum solution without recomputation, (4) providing a built-in sensitivity whereby the effect of variation in the weightings could be evaluated and (5) where the minimum sum of squares criteria could be used as a basis for determining optimum weighting of the terms.

In the dynamic programming terminology the stages correspond to the terms which are being combined; therefore, it is necessary to develop the recursive relation of dynamic programming. For example if the actual values are y_i and the individual predicted values for the first term is x_{1i}, then the objective is to select some value θ_1 such that the function $- f_1(\theta) = \sum_{i=1}^{n} (\theta_1 x_{1j} - y_i)^2$ is minimized subject to $0 \le \theta_1 \le \theta$. Using this same notation scheme, the objective for a two-term equation is to determine

$$f_2(\theta) = \min_{0 \le \theta_2 \le \theta} \left[\sum_{i=1}^{n} (\theta_1 x_{1i} + \theta_2 x_{2i} - y_i)^2 \right]$$

$$= \min_{0 \le \theta_2 \le \theta} \left[\sum_{i=1}^{n} (\theta_1^2 x_{1i}^2 + \theta_2^2 x_{2i}^2 + y_i^2 + 2\theta_1 x_{1i} \theta_2 x_{2i} - 2\theta_1 x_{1i} y_i - 2\theta_2 x_{2i} y_i \right]$$

However, $$f_1(\theta) = \min_{0 \le \theta_1 \le \theta} \left[\sum_{i=1}^{n} (\theta_1^2 x_{1i}^2 - 2\theta_1 x_{1i} + y_i^2) \right]$$

and $$\sum_{i=1}^{n} (\theta_2 x_{2i} - y_i)^2 = \sum_{i=1}^{n} (\theta_2^2 x_{2i}^2 - 2\theta_2 x_{2i} y_i + y_i^2)$$

$$\therefore\ f_2(\theta) = \min_{0 \le \theta_2 \le \theta} \left[\sum_{i=1}^{n} (\theta_2 x_{2i} - y_i)^2 - \sum_{i=1}^{n} y_i^2 + 2 \sum_{i=1}^{n} \theta_1 x_{1i} \theta_2 x_{2i} + f_1(\theta - \theta_2) \right].$$

It is probably only of interest at this point, but should be clarified for the next stage of the computation is that the values of $\theta_1 X_{1i}$ are fixed, based upon the value θ_1 takes on in order to optimize $f_1(\theta - \theta_2)$ i.e., some value of $0 \le \theta_1 \le (\theta - \theta_2)$ which minimizes the error sum of squares in stage 1. Therefore, in order to denote the fixed values of the previous stage as being different from the variable values in the cross product term let $z_{2i} = \theta_1 x_{1i}$. It should also be pointed out that $\sum_{i=1}^{n} y_i^2$ is simply a constant.

$$\therefore\ f_2(\theta) = \min_{0 \le \theta_2 \le \theta} \left[\sum_{i=1}^{n} (\theta_2 x_{2i} - y_i)^2 - \sum_{i=1}^{n} y_i^2 + 2 \sum_{i=1}^{n} z_{2i}\, \theta_2 x_{2i} + f_1(\theta - \theta_2) \right].$$

It would be possible to write the general recursive relationship at this point, however, there is a subtle point that should be illustrated. This will be accomplished by considering the third stage or third term to be introduced into the model.

$$f_3(\theta) = \min_{0 \leq \theta_3 \leq \theta} \left[\sum_{i=1}^{n} (\theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i} - y_i)^2 \right]$$

$$= \min_{0 \leq \theta_3 \leq \theta} \left[\sum_{i=1}^{n} (\theta_1^2 x_{1i}^2 + 2\theta_1 x_{1i} \theta_2 x_{2i} + 2\theta_1 x_{1i} \theta_3 x_{3i} \right.$$

$$- 2\theta_1 x_{1i} y_i + \theta_2^2 x_{2i}^2 + 2\theta_2 x_{2i} \theta_3 x_{3i}$$

$$\left. - 2\theta_2 x_{2i} y_i + \theta_3^2 x_{3i}^2 - 2\theta_3 x_{3i} y_i + y_i^2) \right]$$

However, $$f_2(\theta) = \sum_{i=1}^{n} (\theta_1^2 x_{1i}^2 + \theta_2^2 x_{2i}^2 + y_i^2 + 2\theta_{1\;1}\theta_{2\;2}$$

$$- 2\theta_3 x_{3i} y_i + y_i^2)$$

and $$\sum_{i=1}^{n} (\theta_3 x_{3i} - y_i)^2 = \sum_{i=1}^{n} (\theta_3^2 x_{3i}^2 - 2\theta_3 x_{3i} y_i + y_i^2)$$

$$\therefore\quad f_3(\theta) = \min_{0 \leq \theta_3 \leq \theta} \left[\sum_{i=1}^{n} (\theta_3 x_{3i} - y_i)^2 - \sum_{i=1}^{n} y_i^2 \right.$$

$$\left. + 2 \sum_{i=1}^{n} \theta_3 x_{3i} (\theta_1 x_{1i} + \theta_2 x_{2i}) + f_2(\theta - \theta_3) \right]$$

The point that should be noted is that $(\theta_1 x_{1i} + \theta_2 x_{2i})$ is fixed for a given $(\theta - \theta_3)$ and this is the computed values of the model through the previous stage which minimized the error sum of squares for the specified sum of weights $(\theta_1 + \theta_2)$. Note, $(\theta - \theta_3) = (\theta_1 + \theta_2)$. Therefore, if these fixed values of $(\theta_1 x_{1i} + \theta_2 x_{2i})$ are represented as $z_{3i} = (\theta_1 x_{1i} + \theta_2 x_{2i})$ then;

$$f_3(\theta) = \min_{0 \leq \theta_3 \leq \theta} \left[\sum_{i=1}^{n} (\theta_3 x_{3i} - y_i)^2 - \sum_{i=1}^{n} y_i^2 + 2 \sum_{i=1}^{n} \theta_3 x_{3i} z_{3i} \right.$$

$$\left. + f_2(\theta - \theta_3) \right]$$

This type of dynamic programming formulation requires that n additional values of z_{si} be carried from stage to stage. It does not require that they be retained for all previous stages only the preceding one since they are cumulative in nature. Then

$$z_{si} = \sum_{j=1}^{s} \theta_j x_{ji} \;, \quad i = 1, 2, \ldots, n \,.$$

for the computations of stage s. The n stage recursive relation can be written as:

$$f_n(\theta) = \min_{o \leq \theta_n \leq \theta} \left[\sum_{i=1}^{n} (\theta_n x_{ni} - y_i)^2 - \Sigma y_i^2 + 2 \sum_{i=1}^{n} \theta_n x_{ni} z_{ni} + f_{n-1}(\theta - \theta_n) \right], \quad n \geq 2$$

In order to compute these values by standard methods it will be necessary to make θ discrete. The increments can be refined to any level necessary in order to obtain a satisfactory weighting of the terms. This is essentially a one dimensional allocation problem with minor modifications.

Normally it would be expected that $\sum_{j=1}^{n} \theta_j = 1$ would be the constraint on θ; however, due to the built-in sensitivity of dynamic programming it may be desirable to constrain θ to a sum greater than 1 in order to determine if certain downward biases may be contained in the original individual terms. The dynamic programming solution will provide solutions for all values of θ. Further, it will also provide solutions for all arrangements of consecutive groupings of terms with the last term being deleted each time. By ordering the terms according to their suspected importance with respect to the model being constructed, the contribution of each added term can be considered for deletion in the reverse order in which it entered into the calculations.

A computer program can be provided for this algorithm if it is considered to be of more than passing interest.

PART V

PREDICTION OF RUN-OUT COSTS

By

Sidney P. Brown

For

Industrial Engineering Department

Texas A&M University

May 1966

INTRODUCTION

Various methods have been used to estimate the total cost of space programs at various points during the life of the program. Due to various changes that are a normal part of major programs, the estimates change with time and the amount of information available. It is generally considered that the total program cost estimates approach the actual cost as the program nears the end. There have been numerous studies to determine the reasons and appropriate adjustments that could be made to estimate costs. This study was limited to the utilization of Gemini data for the extrapolation of Apollo data. However, the general problem of selecting a cost distribution function over time from among many as a best fit to a partial cost distribution, and the use of the one selected to extrapolate the partial costs to a completed cost was considered.

The hypothesis was that a subsystem of a completed project would follow the percent time vs. percent cost curve of some subsystem of a previously completed project. The subsystems of the uncompleted project would not necessarily follow the curve of the same subsystem of the completed project because of such factors as the amount of parallel development, technological diffculties, program changes and other similar reasons. Therefore, the choice of the best fitting curve should be made from among a "population" of significantly different curves of subsystems of completed projects. Then the curve which approximates the available data the closest should be used to estimate the run-out cost.

There are three basic phases to the estimation of run-out costs in this method:

(1) The determination of a polynomial to fit each "population" curve.

(2) The determination of the best fitting curve among the population

to the uncompleted subsystem data and

(3) The determination of the run-out costs. These phases are shown on the flow chart of Figure 1.

DETERMINE POLYNOMIAL

FOR EACH OF THE SIGNIFICANTLY DIFFERENT COMPLETED SUBSYSTEM CURVES, DETERMINE A THIRD DEGREE POLYNOMIAL THAT BEST FITS THE AVAILABLE DATA POINTS

↓

DETERMINE BEST FITTING CURVE

FOR EACH UNCOMPLETED SUBSYSTEM, DETERMINE THE BEST FIT OF THE AVAILABLE DATA POINTS TO EACH COMPLETED SUBSYSTEM CURVE AND CHOOSE THE ONE THAT BEST FITS

↓

DETERMINE RUN-OUT COST

USING THE BEST FITTING CURVE, ANY COST DATA POINT CAN BE USED TO DETERMINE THE RUN-OUT COST.

FIGURE 1. FLOW CHART OF RUN-OUT COST ESTIMATION PROCESS

DETERMINE POLYNOMIAL

The data of a completed project which was available to Texas A&M was the percent time vs. percient cost curves of Gemini. This data consisted of four intermediate points through which a smooth curve had been drawn (NASA/MSC data).

To obtain an equation of the different curves, a third degree polynomial was determined to be the best general fit to the curves. The general form of the equation is:

$$f(x) = AX + BX^2 + CX^3$$

To determine the coefficients of this equation, the method of least squares was used. A set of four simultaneous normal equations, involving summations of the decimal percent time raised to powers from zero to six and the same summations except being multiplied by percent cost with the powers ranging from zero to three, is solved to give the desired coefficients. The more data points that are used in the computations, the better the fit is expected to be. The degree of the polynomial could be increased if desired in order to obtain a still better fit; however, due to the nature of the real world problem, tests should be made in order to avoid inconsistent changes in the value of the tangent.

This technique was used to determine the equation and was found to give satisfactory results.

A method of determining an equation for all the curves had been found, but now the problem of which curves were significatnly different needed to be resolved. Since the curves go through the points (0,0) and (1,1), the curves can be considered as cumulative distribution curves. Therefore, it is possible to differentiate third degree polynomials to obtain quadratic density functions.

If it is assumed that the third degree (or any degree for that matter) adequately describes the distribution function, then by using the method of moments a Beta density can be fitted to the polynomial. For example, if the cubic is obtained of the form $Y = AX + BX^2 + CX^3$ which will be the case if the curve is forced through (0,0) and (1,1) then the derivative is

$$Y' = A + 2BX + 3CX^2 \quad .$$

(Note: A, B and C are known constants determined by the least squares fit). The first moment will be

$$\begin{aligned} E(x) &= \int_0^1 (AX + 2BX^2 + 3CX^3)\, dx \\ &= \left. \frac{AX^2}{2} + \frac{2BX^3}{3} + \frac{3CX^3}{4} \right|_0^1 \\ &= \frac{A}{2} + \frac{2B}{3} + \frac{3C}{4} \end{aligned}$$

Similarly

$$\begin{aligned} E(x^2) &= \int_0^1 (AX^2 + 2BX^3 + 3CX^4)\, dx \\ &= \frac{A}{3} + \frac{2B}{4} + \frac{3C}{5} \end{aligned}$$

The corresponding first and second moments about the origin for the Beta density can be shown to be $\frac{\alpha + 1}{\alpha + \beta + 2}$ and $\frac{(\alpha + 2)\ (\alpha + 1)}{(\alpha + \beta + 3)(\alpha + \beta + 2)}$ respectively. Then by solving two equations for two unknowns, the two parameters of the Beta density (which uniquely describes the Beta) may be determined.

$$\frac{\alpha + 1}{\alpha + \beta + 2} = \frac{A}{2} + \frac{2B}{3} + \frac{3C}{4} \quad \text{and} \quad \frac{(\alpha + 2)\ (\alpha + 1)}{(\alpha + \beta + 3)\ (\alpha + \beta + 2)} = \frac{A}{3} + \frac{2B}{4} + \frac{3C}{5}$$

However, it should be pointed out that the solution of this system of equations will result in two pairs of α's and β's. Hopefully, one pair will be infeasible due to the restriction on α and β as parameters of the Beta density i.e., $\alpha, \beta > -1$. If this is not the case, then the analyst must make a decision as to which function best describes his data.

However, it should be pointed out that the first derivative of a polynomial that passes through (0,0) and (1,1) will qualify as a probability density function. It is of no particular consequence, except the connection with another section of this report is that the sum of the coefficients of the polynomial used as the distribution function will be equal to unity. In terms of the example given above, A + B + C = 1.

For the specific problem of runout costs of Apollo, the Beta density is not a critical area; however, there was an additional problem of distinguishing which of the cost distributions are significantly different. This is not a particularly difficult problem if the number of samples and conditions of independence in these samples are met; however, due to the nature of the data this is not a straightforward application of statistical inference type of test of significance. Research is continuing in this area.

DETERMINE BEST FITTING GEMINI CURVE FOR APOLLO DATA

The data available for an uncompleted subsystem is given as a cost at a certain time (NASA/MSC data). From the data, the percent time of the project is known since the project length is known. What is needed is the percent cost each of the points represent.

The data points must be tested against each significantly different completed system curve in such a way as to get the best fit and then choose the one curve which gives the best fit.

Since the ratios of the cost data points are known, it is possible to place the first intermediate point at a percent cost of X. With n intermediate data points of an uncompleted subsystem, the n-1 remaining data points are at a height k_2X, $k_3X, \ldots, k_nX$ (see Figure 2).

Using the method of least squares to provide the best fit,

$$(1) \quad (\bar{Y}_1-X)^2 + (\bar{Y}_2-k_2X)^2 + \ldots + (\bar{Y}_n-k_nX)^2 = \text{minimum}$$

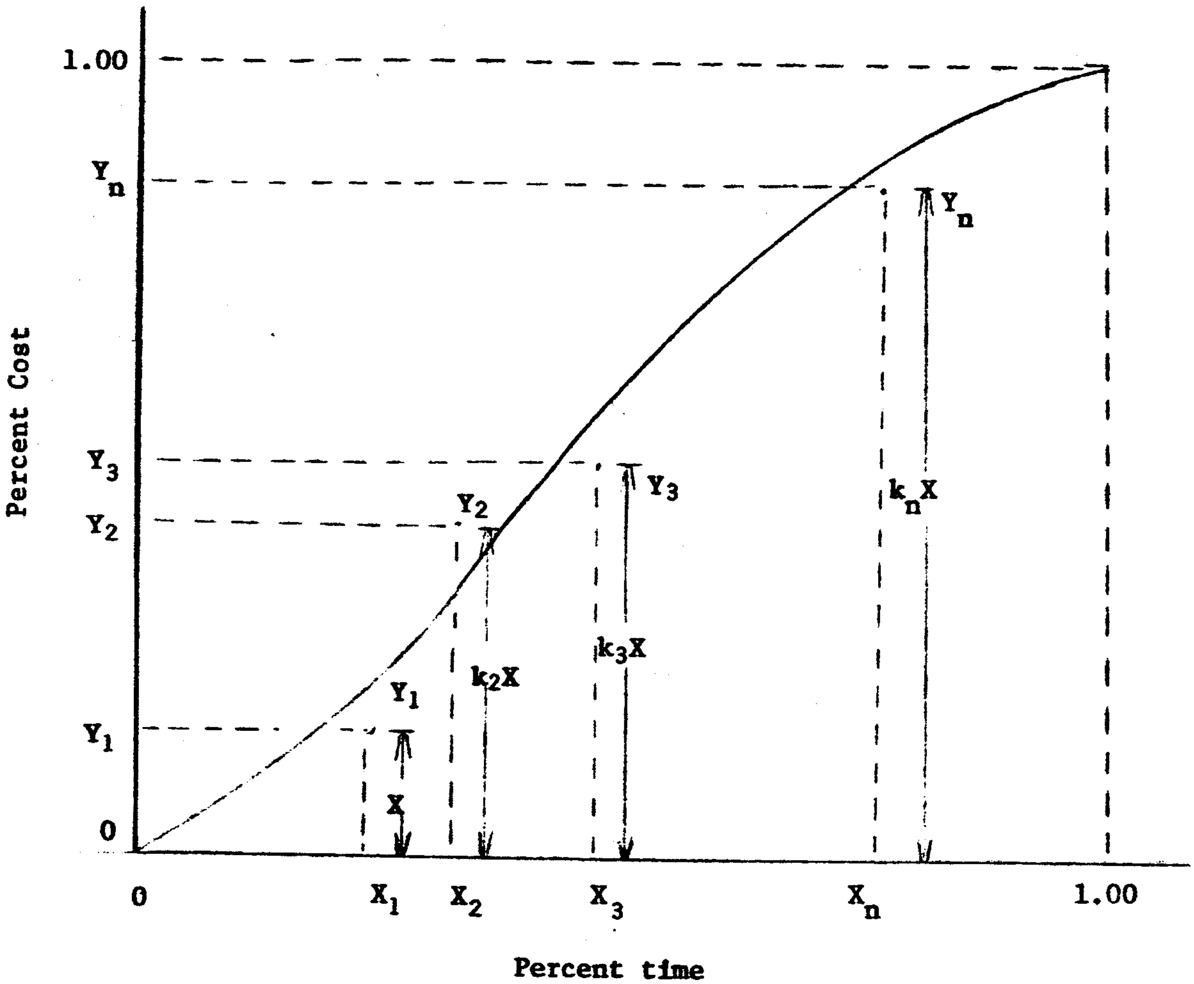


Figure 2. Method of Cost Ratio Determination

Where Y_1, Y_2, ..., Y_n are the values of the percent cost of the completed subsystem at the percent time of the data points of the uncompleted subsystem, these values are obtained from the derived polynomials of the completed subsystem.

Taking the derivative of (1) and setting it equal to zero to obtain the minimum, (2) $2(Y_1-X)(-1) + 2(Y_2 - k_2X)(-k_2) + 2(Y_3 - k_3X)(-k_3) + \ldots + 2(Y_n - k_nX)(-k_n) = 0$

Solving for X;

$$(3) \quad X = \frac{Y_1 + k_2Y_2 + k_3Y_3 + \ldots + k_nY_n}{1 + k_2^2 + k_3^2 + \ldots + k_n^2}$$

By knowing the value of X, the value of the sum of the squares of (1) can be determined to obtain a measure of fit for the uncompleted subsystem to one of the completed subsystem curves. Repeating this with each of the completed subsystem curves gives a measure of fit for each of these curves.

By using this measure of fit, the best fitting curve can be determined by choosing the curve which had a minimum value for the sum of the squares.

DETERMINE RUN-OUT COST

With the best fitting completed subsystem curve chosen, and the uncompleted sybsystem data points located, the run-out cost can be estimated.

The method used to choose the best curve placed the data points of the uncompleted subsystem in the appropriate perspective to the completed subsystem, but each of these points were converted to a percent cost; therefore, each data point is a percentage of the run-out cost, and any one may be used to obtain a projection of the 100% or run-out cost of that subsystem.

Since any point may be used to estimate the run-out cost, the first point will be chosen for convenience since equation (3) located the first intermediate cost point at a percentage of the run-out cost, the relationship of the cost associated with that point to the projected cost is known. Thus, the run-out cost is the cost at the first point divided by the decimal percent of run-out cost (X).

A program which does the complete analysis of cost run-out has been completed. The data of the completed significantly different subsystem and the uncompleted subsystem is the only required information, with the estimated run-out cost as the information provided the user. Program decks will be made available to MSC.

SAMPLE DATA INPUT LISTING

```
$DATA
     3     6     4     1     1     1     1     1     1     1
        INSTRUMENTATION
0.0         0.0         0.285       0.324       0.480       0.600       0.580       0.730
0.675       0.824       1.0         1.0
   STABILIZATION AND CONTROL
0.0         0.0         0.285       0.370       0.480       0.610       0.580       0.720
0.675       0.810       1.0         1.0
        STABILITY CONTROL
0.0         0.0         0.285       0.215       0.480       0.409       0.580       0.518
0.675       0.646       1.0         1.0
          CREW SYSTEMS
0.19        5096.0      0.28        12066.0     0.34        16004.0     0.44        21086.0
   STABILIZATION AND CONTROL
0.19        19889.0     0.28        38176.0     0.34        45689.0     0.44        50852.0
        INSTRUMENTATION
0.19        3343.0      0.28        4510.0      0.34        6896.0      0.44        10717.0
```

```
      DIMENSION P(4,5), X(101), Y(101), Z(101),D(101),  C(101), YA(101),
     1YRO(101),IPIV(4),R1(100),R2(100),R3(100),RATIO(101),SLS(101),
     2W(100,101),NAME(101,5),NAMA(5)
   66 READ (5,13) K,L,M,K1,K2,K3,K4,K5,K6,K7
   13 FORMAT (10I5)
      WRITE (6,33)
   33 FORMAT (1H1,56X,20HCOMPLETED SUBSYSTEMS,//)
      DO 30 J=1,K
      IF (J-1)141,141,142
  142 WRITE (6,143)
  143 FORMAT (1H1)
  141 IF (1-K1) 14,15,14
   15 READ (5,16) (NAME (J,I),I= 1,5)
   16 FORMAT (5A6)
      WRITE (6,17) J, (NAME (J,I),I=1,5)
   17 FORMAT (//,64X,1H(,I3,1H),/,51X,5A6)
   14 P(1,1) = L
      P(1,2)=0
      P(1,3)= 0
      P(1,4) =0
      P(2,4)= 0
      P(3,4)= 0
      P(4,4)= 0
      P(1,5)= 0
      P(2,5)= 0
      P(3,5)= 0
      P(4,5)= 0
      READ (5,12)(X(I), Y(I),I=1,L)
   12 FORMAT (8F10.0)
      DO 50 I=1,L
      P(1,2) = P(1,2) +X(I)
      P(1,3)= P(1,3) + X(I)**2
      P(1,4) = P(1,4) + X(I)**3
      P(2,1) = P(1,2)
      P(2,2) = P(1,3)
      P(2,3) = P(1,4)
      P(2,4) = P(2,4) + X(I)**4
      P(3,1) = P(2,2)
      P(3,2) = P(2,3)
      P(3,3) = P(2,4)
      P(3,4) = P(3,4) + X(I)**5
```

```
      P(4,1) = P(3,2)
      P(4,2) = P(3,3)
      P(4,3) = P(3,4)
      P(4,4) = P(4,4) + X(I)**6
      P(1,5) = P(1,5) + Y(I)
      P(2,5) = P(2,5) + Y(I)*X(I)
      P(3,5) = P(3,5) + Y(I)*X(I)**2
   50 P(4,5) = P(4,5) + Y(I)*X(I)**3
      IF (1-K2) 81,80,81
   80 WRITE (6,36)
   36 FORMAT (//,23X,86HTHE SET OF SIMULTANEOUS EQUATIONS THAT GIVES THE
     1 COEFFICIENTS OF THE COST EQUATION IS,)
      DO 38 I=1,4
      WRITE (6,37) P(I,1),P(I,2),P(I,3), P(I,4), P(I,5)
   37 FORMAT (//,5X,E15.8,5H A + ,E15.8,5H B + ,E15.8,5H C + ,E15.8,
     25H D = , E15.8)
   38 CONTINUE
   81 DET = SIMEQN(P,IPIV,4,4,1.0E-15)
      R1(J) = P(2,5)
      R2(J) = P(3,5)
      R3(J) = P(4,5)
      IF (1-K3) 83,82,83
   82 WRITE (6,51) P(2,5),P(3,5), P(4,5)
   51 FORMAT (//,49X,34HTHE EQUATION OF THE COST CURVE IS,,/,27X,
     3 6HYA = (,E15.8,8H)(X) + (,E15.8,11H)(X**2) + (,F15.8,7H)(X**3))
   83 IF (1-K4) 30,84,30
   84 DO 75 I=1,L
      YA(I) = P(2,5)*X(I) + P(3,5)*X(I)**2 + P(4,5)*X(I)**3
      WRITE (6,55) I, YA(I), I, Y(I)
   55 FORMAT (//,57X,3HYA(,I3,3H) =, F8.4,/,57X,3HY (,I3,3H) =,F8.4)
   75 CONTINUE
   30 CONTINUE
      WRITE (6,26)
   26 FORMAT (1H1,55X,22HUNCOMPLETED SUBSYSTEMS,//)
   59 IF(1-K5)58,111,58
  111 READ (5,108) NAMA
  108 FORMAT (5A6)
      WRITE (6,96) NAMA
   96 FORMAT (//,51X,5A6)
   58 READ (5,18)(Z(I), C(I),I=1,M)
   18 FORMAT (8F10.0)
      DO 31 J=1,K
      DO 25 I=1,M
      W(J,I) = R1(J)*Z(I) + R2(J)*Z(I)**2 + R3(J)*Z(I)**3
```

```
  5 CONTINUE
    RA=0
    RAT = 0
    DO 64 I=1,M
    D(I) = C(I)/C(1)
    RA = RA + D(I)*W(J,I)
    RAT = RAT + D(I)**2
 64 CONTINUE
    RATIO(J) = RA/RAT
    SLS(J) = 0
 91 DO 61 I= 1,M
    SLS(J) = SLS(J) + (W(J,I) -D(I)*RATIO(J))**2
 61 CONTINUE
 31 CONTINUE
    IF (1-K6) 34,90,34
 90 WRITE (6,45)
 45 FORMAT (//,34X,63HTHE SUM OF LEAST SQUARES USING THE I-TH COMPLETE
   1D SUBSYSTEM IS,)
    DO 32 J=1,K
 92 WRITE (6,44) J,SLS(J)
 44 FORMAT (//,53X,4HSLS(,I3,3H) =,E15.8)
 32 CONTINUE
 34 XHI = SLS(1)
    IB = 1
    DO 101 I=2,K
    IF (XHI.LE.SLS(I)) GO TO 101
    XHI = SLS(I)
    IB = I
101 CONTINUE
    IF (1-K7) 93,22,93
 22 WRITE (6,102) IB,(NAME(IB,I),I=1,5)
102 FORMAT (//,44X, 39HTHE CURVE WHICH BEST FITS THE DATA IS (,I3,2H),
   2/, 51X, 5A6)
 93 ROC = C(1 )/RATIO(IB)
    WRITE (6,103) ROC
103 FORMAT (//,49X,21HTHE RUN-OUT COST IS $,F12.1)
    GO TO 59
    END
```

(1)
INSTRUMENTATICN

THE SET OF SIMULTANEOUS EQUATIONS THAT GIVES THE COEFFICIENTS OF THE COST EQUATION IS,

0.60000000E 01 A + 0.30200000E 01 B + 0.21036500E 01 C + 0.16363999E 01 D = 0.34780000E 01

0.30200000E 01 A + 0.21036500E 01 B + 0.16363999E 01 C + 0.13804407E 01 D = 0.23599400E 01

0.21036500E 01 A + 0.16363999E 01 B + 0.13804407E 01 C + 0.12331224E 01 D = 0.17855638E 01

0.16363999E 01 A + 0.13804407E 01 B + 0.12331224E 01 C + 0.11454203E 01 D = 0.14697059E 01

THE EQUATION OF THE COST CURVE IS,
YA = (0.85147243E 00)(X) + (0.14301639E 01)(X**2) + (-0.12815837E 01)(X**3)

YA(1) = 0.
Y (1) = 0.

YA(2) = 0.3292
Y (2) = 0.3240

YA(3) = 0.5965
Y (3) = 0.6000

YA(4) = 0.7249
Y (4) = 0.7300

YA(5) = 0.8322
Y (5) = 0.8240

YA(6) = 1.0001
Y (6) = 1.0000

(2)
STABILIZATION AND CONTROL

THE SET OF SIMULTANEOUS EQUATIONS THAT GIVES THE COEFFICIENTS OF THE COST EQUATION IS,

```
0.60000000E 01 A +  0.30200000E 01 B +  0.21036500E 01 C +  0.16363999E 01 D =  0.35100000E 01

0.30200000E 01 A +  0.21036500E 01 B +  0.16363999E 01 C +  0.13804407E 01 D =  0.23626000E 01

0.21036500E 01 A +  0.16363999E 01 B +  0.13804407E 01 C +  0.12331224E 01 D =  0.17818615E 01

0.16363999E 01 A +  0.13804407E 01 B +  0.12331224E 01 C +  0.11454203E 01 D =  0.14656199E 01
```

THE EQUATION OF THE COST CURVE IS,

```
YA = ( 0.12734183E 01)(X) + ( 0.24018385E 00)(X**2) + (-0.51360290E 00)(X**3)

YA(  1) =  0.
Y (  1) =  0.

YA(  2) =  0.3705
Y (  2) =  0.3700

YA(  3) =  0.6098
Y (  3) =  0.6100

YA(  4) =  0.7192
Y (  4) =  0.7200

YA(  5) =  0.8110
Y (  5) =  0.8100

YA(  6) =  1.0000
Y (  6) =  1.0000
```

UNCOMPLETED SUBSYSTEMS

CREW SYSTEMS

THE SUM OF LEAST SQUARES USING THE I-TH COMPLETED SUBSYSTEM IS,

SLS(1) = 0.54172525E-02

SLS(2) = 0.12370633E-01

SLS(3) = 0.16291000E-02

THE CURVE WHICH BEST FITS THE DATA IS (3),
STABILITY CONTROL

THE RUN-OUT COST IS $ 57865.2

STABILIZATION AND CONTROL

THE SUM OF LEAST SQUARES USING THE I-TH COMPLETED SUBSYSTEM IS,

SLS(1) = 0.62256618E-02

SLS(2) = 0.50580048E-02

SLS(3) = 0.41257425E-02

THE CURVE WHICH BEST FITS THE DATA IS (3),
STABILITY CONTROL

THE RUN-OUT COST IS $ 159041.7

INSTRUMENTATION

THE SUM OF LEAST SQUARES USING THE I-TH COMPLETED SUBSYSTEM IS,

SLS(1) = 0.87812680E-02

SLS(2) = 0.17973664E-01

(3)
STABILITY CONTROL

THE SET OF SIMULTANEOUS EQUATIONS THAT GIVES THE COEFFICIENTS OF THE COST EQUATION IS,

```
0.60000000E 01 A +  0.30200000E 01 B +  0.21036500E 01 C +  0.16363999E 01 D =  0.27880000E 01

0.30200000E 01 A +  0.21036500E 01 B +  0.16363999E 01 C +  0.13804407E 01 D =  0.19940850E 01

0.21036500E 01 A +  0.16363999E 01 B +  0.13804407E 01 C +  0.12331224E 01 D =  0.15802859E 01

0.16363999E 01 A +  0.13804407E 01 B +  0.12331224E 01 C +  0.11454203E 01 D =  0.13499525E 01
```

THE EQUATION OF THE COST CURVE IS,

```
YA = ( 0.48845712E 00)(X) + ( 0.10070867E 01)(X**2) + (-0.49555767E 00)(X**3)
```

```
YA(  1) =  0.
Y (  1) =  0.

YA(  2) =  0.2095
Y (  2) =  0.2150

YA(  3) =  0.4117
Y (  3) =  0.4090

YA(  4) =  0.5254
Y (  4) =  0.5180

YA(  5) =  0.6362
Y (  5) =  0.6460

YA(  6) =  1.0000
Y (  6) =  1.0000
```

```
SLS(  3) = 0.23101872E-02

THE CURVE WHICH BEST FITS THE CATA IS (  3),
STABILITY CONTROL

THE RUN-OUT CCST IS $      27319.2
```

PART VI

QUANTIFICATION OF EXPERTISE

By

Grady C. Haynes

For

Industrial Engineering Department

Texas A&M University

May 1966

QUANTIFICATION OF EXPERTISE

The objective of this research is the development of a reliable method of predicting various futuristic cost functions. The data to be used will consist of subjective evaluations in the form of 'expert' opinion, 'expert' being defined as an individual whose answers to certain questions can be considered reliable due to his experience in the area in question. A computer program will be developed to facilitate the quantification of expertise and to provide parametric type data upon which decisions can be based.

Some work in this area has previously been done in this area by the Rand Corporation. Project Delphi, of the early 1950's, was an attempt to provide answers to questions pertaining to the ability of the United States to withstand a nuclear attack. The experiment tried to provide the answers to these questions, but the data gathered was not subjected to any extensive statistical analysis. At the present time calculations based upon the multinomial distribution are being performed, with these preliminary calculations giving favorable results.

Other work which has been done in this area is the PATTERN (Minneapolis-Honeywell) program which was developed by Honeywell for the Department of Defense and subsequently used by NASA. This program utilizes expert opinion to rank various space programs according to their value from the standpoint of technological advance, national prestige, etc.

The research which is to be undertaken differs from these two projects (Delphi and Pattern) in that it will include the statistical analysis not undertaken in Delphi, and is not a ranking method as is Pattern.

The problem to which the research is currently being applied is one of determining percent cost/percent time curves for various NASA programs. The 'experts' being questioned are located aa Marshall Space Flight Center, Huntsville, Alabama; and the Manned Spacecraft Center, Houston, Texas.

The progress to the present has consisted of work on the preparation of a statistical model, upon which will be based the statistical analysis of the data received. The submission of the first set of questionnaires to MSFC Huntsville and to MSC Houston has also been accomplished. The first set of forms asks the experts involved to sketch their ideas on the various cost-time relationships.

The data received from the initial questionnaires will be used in several ways. The reaction of the experts to the questionnaire is of much interest. From their reactions, it is hoped that the number of questions to be included in future questionnaires can be determined so as to yield as much information as possible while at the same time not being so lengthy that the interest of the experts is lost. Also, the data returned will provide information on how many curves should be used in the second set of questionnaires and the specific types of curves to be used.

The second questionnaire will differ from the first in that pre-drawn curves will be sent to the experts. The experts will be asked to choose which of the curves, in their opinion, best describes each specific cost-time relationship. Only one curve can be chosen for each category. The experts will also be asked for a subjective rating of their confidence in their own answers, and these ratings will be used to weight the data before analysis. In this way, the opinion of a judge who is more experienced in a certain area would receive more consideration than the opinion of judge with less experience.

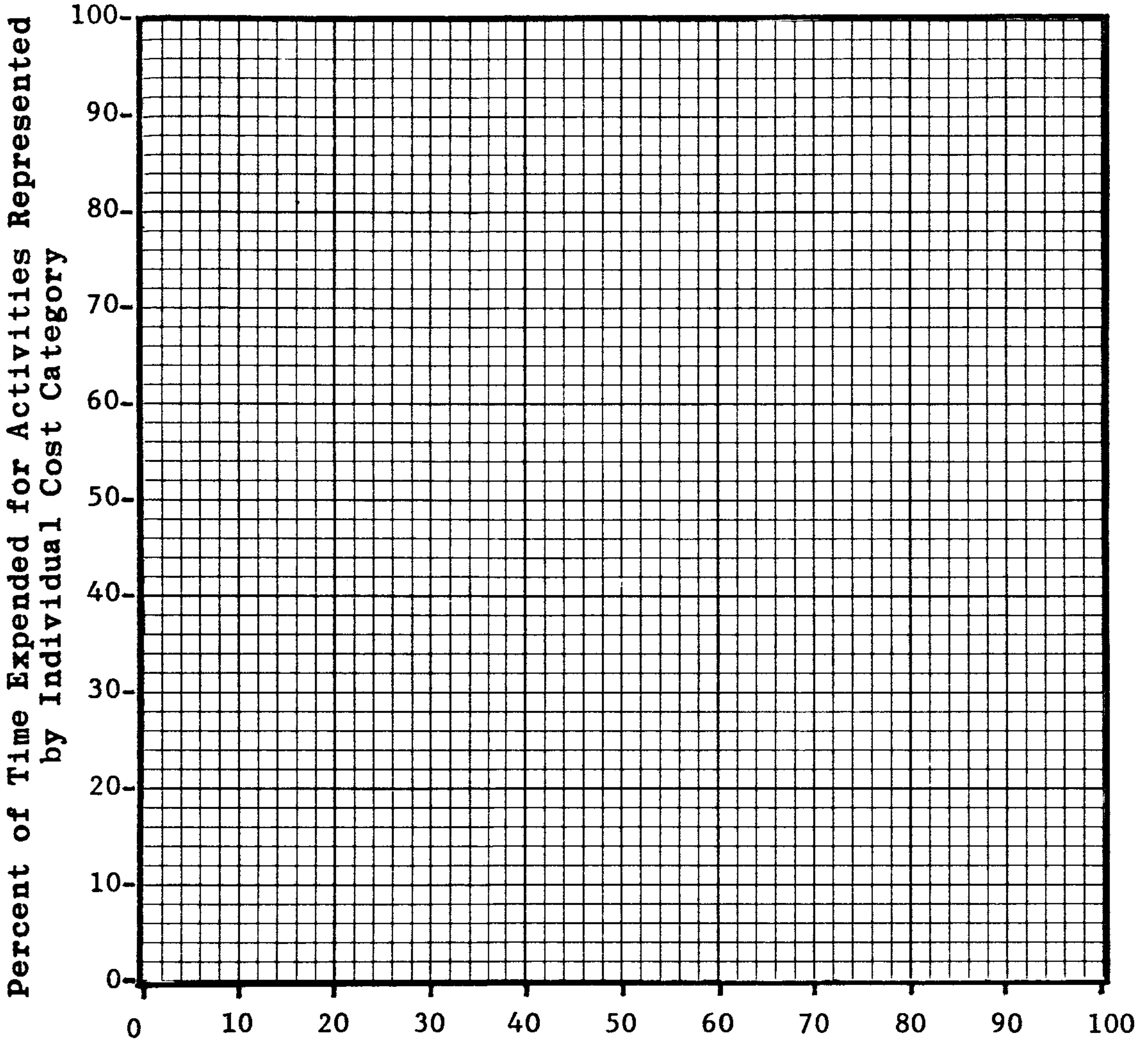
At this level it is hoped that statistical analysis will give some indication of the convergence or divergence of the expert opinions. In the cases where this analysis shows the judges in general agreement or the curve which best describes a specific cost-time relationship, it will be assumed that the appropriate area is quantified and the cost category eliminated from further study and data collection. For those categories where the judges do not agree, questionnaires will again be sent, this time giving each expert additional information which might be of some help in re-evaluating his first answer.

This procedure may still fail to bring the judges into agreement on which curve best describes each specific category. The opinions may, however, group around two or at most three values. In this case, it may be concluded that, for this specific category, the curves are indistinguishable. If one curve is needed, a new curve could conceivably be generated which would be a weighted combination of the curves in the majority.

If the judges disagree widely after several questionnaires, it can only be concluded that the judges have different experience in the area, thereby biasing the results.

It is hoped that this research can contribute something to the area of quantification of subjective information. The results of the first questionnaires indicate a reasonably good and consistent response from the experts contacted. The form of the questionnaire submitted is an addendum to this section of the report. Twenty blank graph forms were submitted with the list of cost categories. These forms will have pre-drawn curves on the next level of questioning and the expert will simply record the number of the cost category for the applicable pre-drawn curve.

Curve Number ____



Percent of Total Cost Expended in Specified Cost Category(s)

This curve is applicable for cost categories:

____, ____, ____, ____, ____, ____, ____, ____, ____,

____, ____, ____, ____, ____, ____, ____, ____, ____.

SUBSYSTEM LEVEL

1. PRIMARY STRUCTURE - DEVELOPMENT ENGINEERING
2. PRIMARY STRUCTURE - INITIAL TOOLING
3. PRIMARY STRUCTURE - TEST ARTICLES
4. PRIMARY STRUCTURE - SUSTAINING ENGINEERING
5. PRIMARY STRUCTURE - MANUFACTURING
6. SERVICE STRUCTURE - DEVELOPMENT ENGINEERING
7. SERVICE STRUCTURE - INITIAL TOOLING
8. SERVICE STRUCTURE - TEST ARTICLES
9. SERVICE STRUCTURE - SUSTAINING ENGINEERING
10. SERVICE STRUCTURE - MANUFACTURING
11. PROPULSION - DEVELOPMENT ENGINEERING
12. PROPULSION - INITIAL TOOLING
13. PROPULSION - TEST ARTICLES
14. PROPULSION - SUSTAINING ENGINEERING
15. PROPULSION - MANUFACTURING
16. STABILITY AND CONTROL - DEVELOPMENT ENGINEERING
17. STABILITY AND CONTROL - INITIAL TOOLING
18. STABILITY AND CONTROL - TEST ARTICLES
19. STABILITY AND CONTROL - SUSTAINING ENGINEERING
20. STABILITY AND CONTROL - MANUFACTURING
21. REACTION CONTROL - DEVELOPMENT ENGINEERING
22. REACTION CONTROL - INITIAL TOOLING
23. REACTION CONTROL - TEST ARTICLES
24. REACTION CONTORL - SUSTAINING ENGINEERING
25. REACTION CONTROL - MANUFACTURING
26. GUIDANCE AND NAVIGATION - DEVELOPMENT ENGINEERING
27. GUIDANCE AND NAVIGATION - INITIAL TOOLING
28. GUIDANCE AND NAVIGATION - TEST ARTICLES
29. GUIDANCE AND NAVIGATION - SUSTAINING ENGINEERING
30. GUIDANCE AND NAVIGATION - MANUFACTURING

31. ELECTRICAL POWER (FUEL CELLS) - DEVELOPMENT ENGINEERING
32. ELECTRICAL POWER (FUEL CELLS) - INITIAL TOOLING
33. ELECTRICAL POWER (FUEL CELLS) - TEST ARTICLES
34. ELECTRICAL POWER (FUEL CELLS) - SUSTAINING ENGINEERING
35. ELECTRICAL POWER (FUEL CELLS) - MANUFACTURING
36. COMMUNICATIONS - DEVELOPMENT ENGINEERING
37. COMMUNICATIONS - INITIAL TOOLING
38. COMMUNICATIONS - TEST ARTICLES
39. COMMUNICATIONS - SUSTAINING ENGINEERING
40. COMMUNICATIONS - MANUFACTURING
41. INSTRUMENTATION - DEVELOPMENT ENGINEERING
42. INSTRUMENTATION - INITIAL TOOLING
43. INSTRUMENTATION - SUSTAINING ENGINEERING
44. INSTRUMENTATION - MANUFACTURING
45. CREW SYSTEMS - DEVELOPMENT ENGINEERING
46. CREW SYSTEMS - INITIAL TOOLING
47. CREW SYSTEMS - TEST ARTICLES
48. CREW SYSTEMS - SUSTAINING ENGINEERING
49. CREW SYSTEMS - MANUFACTURING
50. LAUNCH ESCAPE (EXTERNAL) - DEVELOPMENT ENGINEERING
51. LAUNCH ESCAPE (EXTERNAL) - INITIAL TOOLING
52. LAUNCH ESCAPE (EXTERNAL) - SUSTAINING ENGINEERING
53. LAUNCH ESCAPE (EXTERNAL) - MANUFACTURING
54. LAUNCH ESCAPE (INTEGRAL) - DEVELOPMENT ENGINEERING
55. LAUNCH ESCAPE (INTEGRAL) - INITIAL TOOLING
56. LAUNCH ESCAPE (INTEGRAL) - SUSTAINING ENGINEERING
57. LAUNCH ESCAPE (INTEGRAL) - MANUFACTURING
58. LANDING AND RECOVERY (ATMOSPHERIC) - DEVELOPMENT ENGINEERING
59. LANDING AND RECOVERY (ATMOSPHERIC) - INITIAL TOOLING
60. LANDING AND RECOVERY (ATMOSPHERIC) - SUSTAINING ENGINEERING
61. LANDING AND RECOVERY (ATMOSPHERIC) - MANUFACTURING

62. LANDING AND RECOVERY (NON-ATMOSPHERIC) - DEVELOPMENT ENGINEERING

63. LANDING AND RECOVERY (NON-ATMOSPHERIC) - INITIAL TOOLING

64. LANDING AND RECOVERY (NON-ATMOSPHERIC) - SUSTAINING ENGINEERING

65. LANDING AND RECOVERY (NON-ATMOSPHERIC) - MANUFACTURING

66. ADAPTER - DEVELOPMENT ENGINEERING

67. ADAPTER - INITIAL TOOLING

68. ADAPTER - TEST ARTICLES

69. ADAPTER - SUSTAINING ENGINEERING

70. ADAPTER - MANUFACTURING

MODULE LEVEL

71. SUBSYSTEM INTEGRATION AND INSTALLATION

72. GROUND TESTING

73. FLIGHT TESTING

74. LAUNCH SITE SUPPORT

75. RECOVERY OPERATIONS

76. MISSION CONTROL

77. MISSION PLANNING AND ANALYSIS

78. FLIGHT CREW OPERATIONS - FLIGHT SIMULATORS AND TRAINERS

79. FLIGHT CREW OPERATIONS - OTHER

80. GROUND SUPPORT EQUIPMENT - CHECKOUT

81. GROUND SUPPORT EQUIPMENT - SERVICE, AUXILIARY, AND HANDLING

82. EXPERIMENTS

83. RECONDITIONING

SPACECRAFT LEVEL

84. SUPPORTING DEVELOPMENT

85. OTHER RESEARCH AND DEVELOPMENT

86. FACILITIES

87. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT

88. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - ACCEPTANCE TEST STANDS

89. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - PAD RELATED FACILITIES

90. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - ASSEMBLY BUILDING AND

CHECKOUT FACILITIES

91. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - OTHER COMPLEX RELATED FACILITIES

92. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - LAUNCH CENTER RELATED AND OTHER MISCELLANEOUS FACILITIES

93. OPERATIONAL LAUNCH FACILITIES AND EQUIPMENT - GROUND SUPPORT EQUIPMENT

94. RECOVERY AND RECONDITIONING FACILITIES

95. ADMINISTRATIVE OPERATIONS

PART VII

PRODUCTION COST MODELS

By

William M. Barnes

For

Industrial Engineering Department

Texas A&M University

May 1966

ABSTRACT

Due to the tremendous growth in the size and complexity of our industrial organizations, a need has arisen for scientific quantitative approaches to the solution of a myriad of complex problems. It has been the practice in the past to solve production cost and efficiency problems by use of the standard cost system, but a new technique for the solution of these problems involves the use of production cost models. This procedure utilizes the electrical engineer's servomechanism theory which expresses the variables of a system in some mathematical form which relates the input to the output of the system.

The models that are most often encountered in a production process include the simple block, recycle case, cleanup case, closed loop case, and the recycle with a primary loss case. These individual blocks can then be combined to simulate a complete system. It is also possible to use this technique for optimizing the profit yielded on a multiproduct system.

TABLE OF CONTENTS

TABLE OF ILLUSTRATIONS

I. INTRODUCTION AND PRELIMINARY DISCUSSION

During recent years the phenomenal growth in size and complexity of our economy and our industrial organizations has brought about an intense need for improved techniques of management and control of production operations. Executives in industry have seen the need for a scientific quantitative approach to solve their problems, problems which have increased to a size such that wrong decisions can be tremendously costly while the information to make decisions has become a function of a myriad of variables, whose relationships to each other are quite complex. Highly complex business or industrial organizations are made of many small units or components, each of the units performing a specific duty which contributes in a small manner to the existence of the organization. When all the components and their interactions are integrated together the result is the organization.

In the past the cost of a product, the efficiency of the system producing the product, and the profits realized on the product were usually calculated using the standard cost system. Standard costs have been defined as "a forecast or predetermination of what costs should be under projected conditions, serving as a basis of cost control and as a measure of productive efficiency when ultimately compared with actual costs. The volume level and the set of circumstances under which the product is produced must be carefully determined and the underlying details spelled out. Until these requirements have been met, no attempt to develop standard costs should be made; in fact, no steps can be taken in that direction because there will not be any standards."

Standard costs were frequently developed after a company or an industry had extensive experience with historical costs. A detailed study and analysis of past costs and the modifications of such costs in the light of current and future conditions provided a background for the introduction of a system of standard costs.

Ordinarily, a simulated system was designed with information regarding station efficiencies obtained by a best estimate of the production manager. Standard times for the labor required on the product were obtained by motion and time analysts performing stop watch time studies. Then, by adding in the cost of material and an allowance for overhead, the projected cost of the product was obtained. The standard cost system made no allowance for the proportion of the labor which had already gone into a product unit which was spoiled further down the production line. Therefore, it seemed necessary to devise a cost control system which would account for the feedback of unacceptable units and the waste of labor performed on spoiled units.

The electrical engineer for years has been utilizing methods which establish mathematical expressions for the components and their interactions within highly complex electronic systems. The concept of "transfer functions" and block diagram flow has been especially useful for solving problems in servomechanism (feedback control) systems. Since these methods have been found to be so highly successful in the analysis of electronic systems, their use may yield fruitful results in analyzing some less technical but just as complex production problems.

Whatever the nature of the many variables in a system, these variables should be able to be expressed in some mathematical form relating them to the system itself. The transfer function is nothing but a mathematical expression or model of the system as a function of its variables, i.e., a mathematical expression of the system in terms of the ratio of the output to the input of the system. This transfer function provides a method whereby the engineer can set down in mathematical expressions a representation of the real system retaining as much as possible its many important characteristics. The expression strives to establish in mathematical terminology most of the pertinent characteristics of the system.

Let us delve into an analysis of a simple electrical circuit as an example of the use of the transfer function by the engineer. Althrough the reader may not be acquainted with the concepts of electricity or electronic circuits, the example itself, nevertheless, is simple enough to show the reader the use of a transfer function. The adoption of the function to industrial and business systems will follow. Consider the electrical circuit shown in Figure 1.

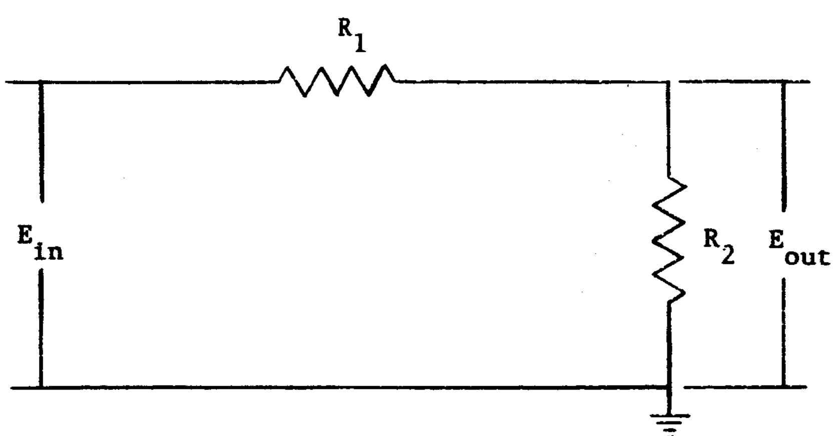


Figure 1

An Electrical Voltage Divider Circuit

This is called a voltage divider circuit and is composed of a resistor R_1 and a resistor R_2. Let us assume that a certain voltage E_{in} is applied to this circuit as shown in the diagram. As the electrical current goes through the components, a certain voltage output isproduced which we call E_{out}. Thus the transformation of the input voltage, E_{in}, to the output voltage, E_{out}, was accomplished by the inherent characteristics of the R_1, R_2 system. E_{out} is different from E_{in} by virtue of the effect R_1 and R_2 had on E_{in}. This effect is written in the transfer function form, $[R_2/(R_1+R_2)]$, which is the mathematical expression of the R_1, R_2 circuit performance on the input E_{in}.

When the input to the circuit is multiplied by the transfer function expression the result is E_{out}. Thus

(input voltage) × (circuit transfer function) = (output voltage)

or

$$(E_{in}) \times [R_2/(R_1+R_2)] = (E_{out})$$

If the value of R_1 and R_2 were 100 and 300 ohms respectively, E_{out} would equal 0.75 E_{in}. Thus, the R_1, R_2 circuit performance transformed the input value of the circuit (E_{in}), to the output value, (0.75 E_{in}).

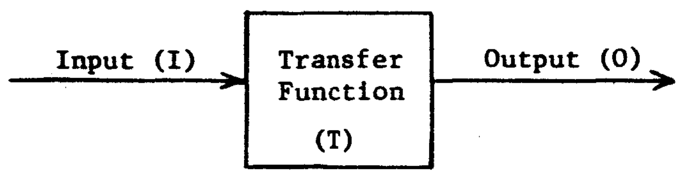
Now, this transfer equation will be written in more general terms so that it may be applicable to any physical system.

(input to system) × (system transfer function) = (output of system)

or

$$(I) \times (T) = (O)$$

To aid in the analysis of overall systems, the engineer utilizes a block diagram technique to indicate the relationship of the above equation (see Figure 2). The function (T) transfers the input (I) to some



$$(I) \times (T) = (0)$$

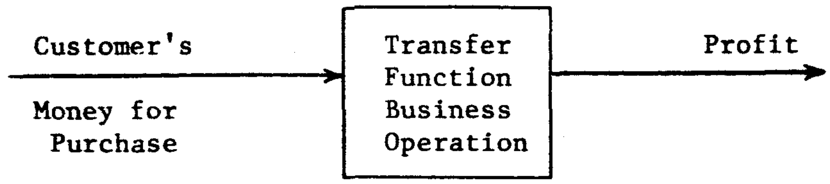
Figure 2

Block Diagram Representation of
a System Transfer Function Equation

output (0). Input (I) is multiplied by the transfer function within the block to obtain the output (0). The use of the block diagram to represent the transfer equation greatly aids in the analysis of complex systems as will be shown later in this paper.

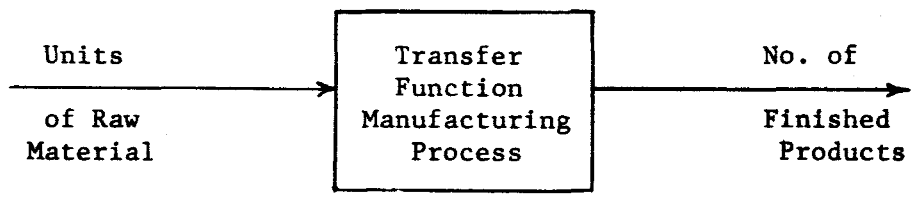
A business operation analysis can be developed also by the transfer function block diagram method. Consider the performance of a business operation to be represented by a mathematical expression or transfer function "T." The transfer function "T" transforms the inputs to the business operation into some output (see Figure 3 for some generalized examples).

Figure 3(a) symbolizes the input to the block as the customer's money received for a purchase. The input is multiplied by the business operation transfer function to obtain the output which is profit. The transfer function in this case could be a function of unit direct cost, overhead cost, taxes, etc. Figure 3(b) symbolizes the input of a certain



(Money) × Transfer function Business Operation = Profit

(a)



Units of Raw Material × Transfer Fucntion Manufacturing Process = Number of Finished Products

(b)

Figure 3

General Transfer Relationships

number of raw materials into a manufacturing process. The raw materials are then transformed into a certain number of finished products by the process. The manufacturing process transfer function may be a function of labor, work materials, efficiency, etc.

At this point it may be of interest to note that the meaningful business ratios that economists and financiers have been monitoring for years as an indication of what a business is doing are comparable to the input-output ratios that the engineers monitor by use of the transfer function to see what electric systems are doing (see Figure 4).

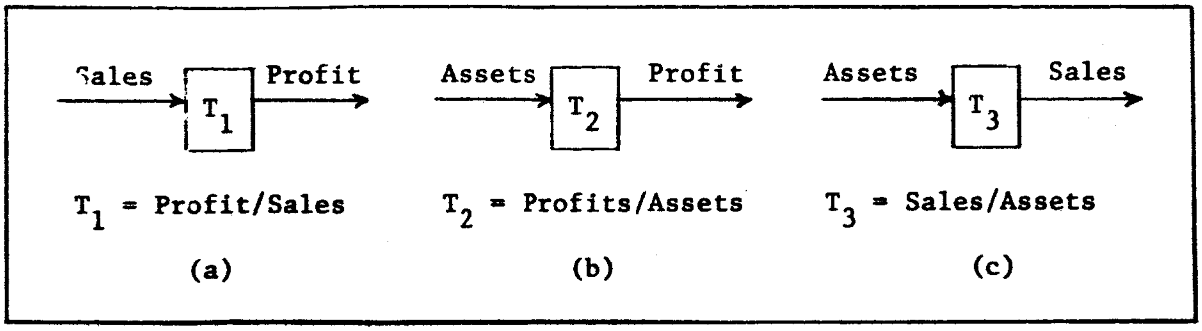


Figure 4

Meaningful Business Ratios
in Transfer Function Terms

Although transfer function techniques can be applied to many business problems, this paper will concentrate upon the use of the transfer function to investigate or solve elementary manufacturing process problems. The true worth of the technique can be seen when investigating complex systems. The use of this technique will be developed through the explanation of basic examples.

II. GENERAL CONSIDERATIONS REGARDING WORK STATION MODELS

This section of the paper will be devoted to the introduction and analysis of the individual work station arrangements that will be most often encountered in a production line process.

Simple Block

The most elementary model is called the "simple block" and consists of a single work station with units of material entering with their associated cost, the application of material and labor at the station, the efficiency of the station which is the probability of sending an acceptable unit on to the next station, and the remainder of the units which have to be scrapped for a salvage value which may be either positive, zero, or even negative, i.e., the company must pay to have the scrap carried away. First the following assumptions are made.

C_o = cost per unit entering process

C_1 = cost added per unit at station

S = salvage value per unit of defective units

p_1 = probability of accepting a unit

K_s = total cost per unit through station

N_I = number of units entering station

N_o = number of good units leaving station

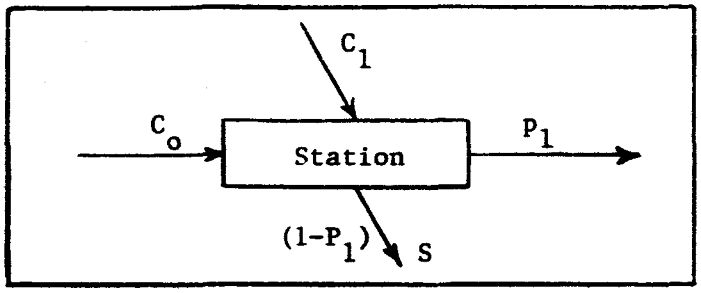


Figure 5

Simple Block

Under the assumptions of this model, it can be seen that the number of units entering the station multiplied by the station efficiency equals the number of acceptable units leaving the station, or

$$N_o = p_1 N_I$$

The number of units which are spoiled and therefore are sold for the salvage value are $(1 - p_1)N_I$. It can be seen then that the cost per unit through the station when 100 units are started is

$$K_s = \frac{100\ C_o + 100\ C_1 - 100(1-p_1)S}{100\ p_1} = \frac{C_o + C_1 - (1-p_1)S}{p_1}$$

A method for computing the expected value and the variance of the cost per good unit through a single work station is shown in Appendix A. The standard deviation of the cost is shown to be $\sigma_K = (C_o + C_1)\sqrt{1-p}/p$. Let's look at an example of a station that has a relatively high probability of outputing a good unit and see just what effect this variance has.

<u>Example</u>: Let p = .99

$$\sigma_K = (C_o + C_1)\frac{\sqrt{1-.99}}{.99} = \frac{(C_o + C_1)(.1)}{.99} = .101(C_o + C_1)$$

The standard deviation of K is approximately 10 percent of $C_o + C_1$.

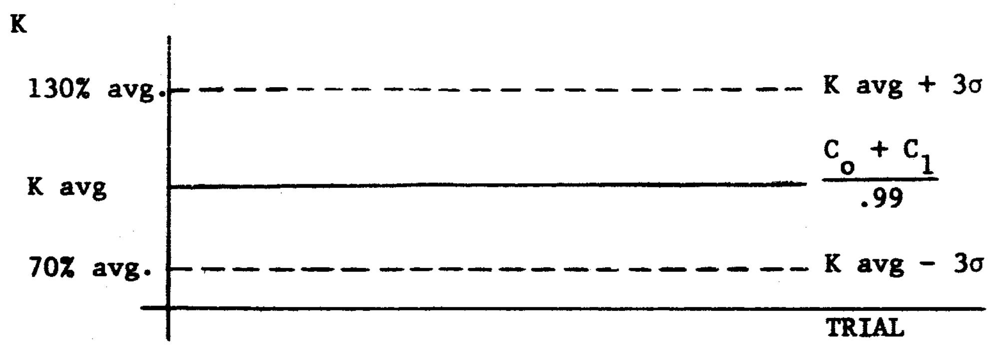


Figure 6

Variance Control Chart

As can be seen from Figure 6, when the process is operating between $\pm$ 3σ the range of possible costs is large.

The variations in the above calculations are not as crucial as they may seem since in an actual industrial operation you are not making one but n units, and since some of the unit costs will be high and others correspondingly low, there will be an averaging out tendency. As n becomes very large the total unit costs will zero in very near the mean.

Simple Blocks in Series

Now let's look at a series of the simple blocks discussed above and determine the total cost per unit through each stage (or station) assuming that we start with 100 units. Also assume zero salvage value for simplicity.

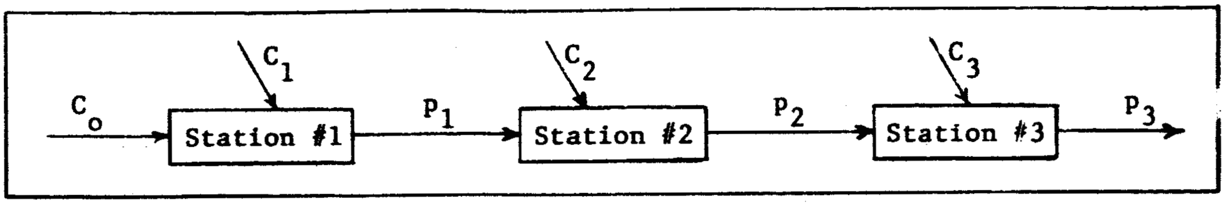


Figure 7

Simple Blocks in Series

$N_o = P_1P_2P_3N_I$

The cost through the first stage is

$$K_1 = \frac{100\ C_o + 100\ C_1}{100\ P_1} = \frac{C_o + C_1}{P_1}$$

The cost through the second stage is

$$K_2 = \frac{100\ K_1 + 100\ C_2}{100\ P_2} = \frac{100[(C_o+C_1)/P_1] + 100\ C_2}{100\ P_2} = \frac{C_o + C_1}{P_1P_2} + \frac{C_2}{P_2}$$

and, finally, the cost through the third stage is

$$K_3 = \frac{100\ K_2 + 100\ C_3}{100\ P_3} = \frac{100[(C_o+C_1)/P_1P_2+C_2/P_2]+100\ C_3}{100\ P_3} = \frac{C_o + C_1}{P_1P_2P_3} + \frac{C_2}{P_2P_3} + \frac{C_3}{P_3}$$

With this average cost per good unit *through* a given stage, one can assess the buildup of cost through the system. In order to determine the cost at an individual stage, one has simply to subtract the cost through the preceding stage from the cost through the stage in question.

Now suppose we decide to attack a certain variable in the system; e.g., the labor and material cost at a certain stage or the percent output of good units at a stage, in order to cut the total cost of production. Which variable should we concentrate on?

Recall that if we have a variable y which is a function of several other variables (x_1, x_2, x_3, ... , x_n) then we have the following relationships:

$$y = f(x_1, x_2, x_3, \ldots, x_n)$$

$$\Delta y = \frac{\partial y}{\partial x_1} \Delta x_1 + \frac{\partial y}{\partial x_2} \Delta x_2 + \ldots + \frac{\partial y}{\partial x_n} \Delta x_n$$

In our example let's look at the cost through stage 2 (K_2) by this analysis:

$$K_2 = \frac{C_o + C_1}{p_1 p_2} + \frac{C_2}{p_2}$$

$$\Delta K_2 = \frac{\partial K_2}{\partial C_o} \Delta C_o + \frac{\partial K_2}{\partial C_1} \Delta C_1 + \frac{\partial K_2}{\partial C_2} \Delta C_2 + \frac{\partial K_2}{\partial p_1} \Delta p_1 + \frac{\partial K_2}{\partial p_2} \Delta p_2$$

$$= \frac{1}{p_1 p_2} \Delta C_o + \frac{1}{p_1 p_2} \Delta C_1 + \frac{1}{p_2} \Delta C_2 - \frac{C_o + C_1}{p_1^2 p_2} \Delta p_1$$

$$- \frac{C_o + C_1}{p_1 p_2^2} \Delta p_2 - \frac{C_2}{p_2^2} \Delta p_2$$

<u>Example</u>: Let $C_o = \$1.00$; $C_1 = \$.50$; $C_2 = \$.10$; $p_1 = .9$; $p_2 = .8$

$$\Delta K_2 = \frac{\Delta C_o}{(.9)(.8)} + \frac{\Delta C_1}{(.9)(.8)} + \frac{\Delta C_2}{(.8)} - \frac{\$1.50 \Delta p_1}{(.9)^2 (.8)} - \frac{\$1.50 \Delta p_2}{(.9)(.8)^2} - \frac{.10 \Delta p_2}{(.8)^2}$$

$$= 1.39 \Delta C_o + 1.39 \Delta C_1 + 1.25 \Delta C_2 - 2.32 \Delta p_1 - 2.60 \Delta p_2 - .16 \Delta p_2$$

$$= 1.39 \Delta C_o + 1.39 \Delta C_1 + 1.25 \Delta C_2 - 2.32 \Delta p_1 - 2.76 \Delta p_2$$

Therefore, it can be seen that a variation in p_2 will cause the greatest corresponding variation in the total unit cost (K_2).

If we could increase p_2 by .1, the resulting savings would be \$.276 per unit. If we could decrease C_o by \$.10, the resulting savings would be \$.139 per unit, etc.

Recycle Case

Thus far only an analysis of a simple system has been made, a type of system the engineer refers to as an "open-loop" system, i.e., a system which has only forward paths or elements. The transfer is always from the input forward to the output. The derivation of a mathematical model of a straightforward system can be a simple task. But now let us look into the derivation of a model for a system described by engineers as a "closed-loop" system, i.e., a system that has a feedback path inherent in its operation. As systems of this type become more complicated with feed-forward and feed-back loops, the normal mathematics becomes more and more burdensome. To analyze such complex systems the engineer, using servomechanism theory, utilizes a method of manipulating block diagrams to investigate the properties of these systems. This manipulation of block diagrams serves to simplify the analysis of a complete system behavior and is used to derive input-output ratios of even very complex systems. To investigate a production closed-loop system, let us look at the manufacturing recycle case. In this system it is assumed that N_I units of raw material are supplied to the manufacturing process and out of manufacturing result some good finished products and some unacceptable or rejected products. If it is assumed that the rejected products can be reconditioned and sent again through the manufacturing process, a simple feedback loop is formed similar to the positive feedback systems in servomechanisms. Thus the rejects are fed back into a reconditioning process so that they can go through the manufacturing process where output is again acceptable and rejected products, and so forth. As can be seen, the reconditioning of the rejected products is not 100 percent; some units are scrapped going through the feedback loop.

Now with the aid of the block diagram technique the recycle case will be investigated. First the following assumptions are made.

C_o = cost per unit entering process

C_1 = cost added per unit during first stage

C_R = cost per unit to rework spoiled units

p_1 = probability of accepting a unit

p_r = probability of accepting a reworked unit

S_R = salvage value per unit of spoiled reworked units

K_R = total cost per good unit

N_I = no. started in system

N_o = no. good units obtained after processing

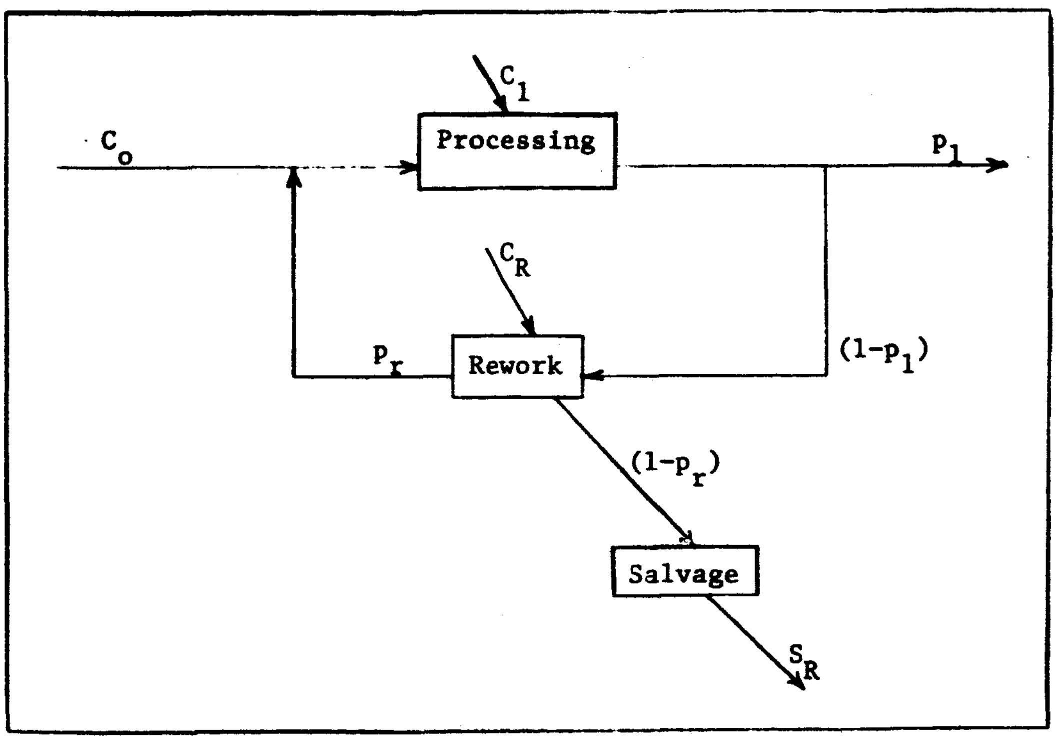


Figure 8

Recycle Case

The total cost per good unit follows:

$$K_R = \frac{C_o + C_1 + (1-p_1)C_R - (1-p_1)(1-p_r)S_R - (1-p_1)p_rC_o}{p_1}$$

The number of good units which will be obtained from the system is

$$N_o = p_1N_I + (1-p_1)p_rp_1N_I + (1-p_1)^2(p_r)^2p_1N_I + \ldots + (1-p_1)^n(p_r)^np_1N_I$$

$$N_o/p_1N_I = [1 + (1-p_1)p_r + (1-p_1)^2(p_r)^2 + \ldots + (1-p_.)^n(p_r)^n]$$

$$N_o/p_1N_I = 1/[1 - (1-p_1)p_r]$$

$$N_o = p_1N_I/[1 - (1-p_1)p_r]$$

The problem that would confront an executive would be that of determining whether or not a reconditioning project would justify the added expense. One aspect that can be investigated to answer this question is the comparative cost per unit calculated by the techniques described.

Recall that the cost per good unit (K_s) under the simple block consideration is

$$K_s = \frac{C_o + C_1 - (1-p_1)S_s}{p_1}$$

Therefore, for reprocessing to be profitable K_s must be greater than K_R, i.e., $K_s > K_R$. Then

$$\frac{C_o + C_1 - (1-p_1)S_s}{p_1} > \frac{C_o + C_1 + (1-p_1)C_R - (1-p_1)(1-p_r)S_R - (1-p_1)p_rC_o}{p_1}$$

$$- S_s > C_R - (1-p_r)S_R - p_rC_o$$

or

$$C_R < (1-p_r)S_R + p_rC_o - S_s$$

$$C_R < S_R - S_Rp_r + p_rC_o - S_s$$

$$\boxed{C_R < p_r(C_o-S_R) + (S_R-S_s)}$$

Therefore, for reprocessing to be profitable, the above equation must hold.

Cleanup Case

Often units are dirty or otherwise undesirable after processing even though they are physically capable of performing their intended job. In this situation there will be a cleanup station which will perform a cleanup operation on the units and then relay them back into the production line.

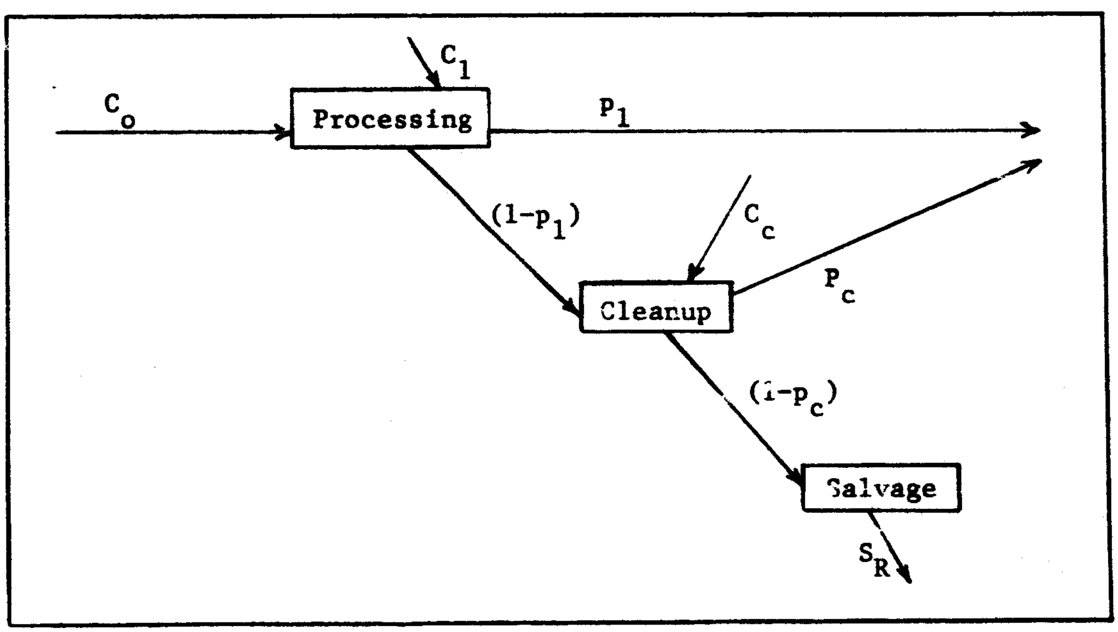


Figure 9

Cleanup Case

$$K_c = \frac{C_o + C_1 + (1-p_1)C_c - (1-p_1)(1-p_c)S_R}{p_1 + (1-p_1)p_c}$$

$$N_o = [p_1 + (1-p_1)p_c]N_I$$

<u>Example</u>: C_o = \$1.00; C_1 = \$.10; C_c = \$.05; S_R = \$.10; p_1 = .90; p_c = .80

<u>Reqd</u>: (a) Cost per good unit (K_c)

(b) Number of units which must be started if 1,000 units are required.

<u>Sol</u>:

(a) $$K_c = \frac{\$1.00 + \$.10 + (1-.9)(\$.05) - (1-.9)(1-.8)(\$.10)}{.9 + (1-.9)(.8)}$$

$$= \frac{\$1.00 + \$.10 + \$.005 - \$.002}{.9 + .08} = \frac{\$1.103}{.98} = \$1.127$$

(b) $1000 = [.9 + (1-.9)(.8)]N_I$

$$N_I = \frac{1000}{.9 + .08} = \frac{1000}{.98} = 1020 \text{ units}$$

Again the problem of justifying the added expense of reconditioning (clean-up in this case) is raised. As before,

$$K_s > K_c$$

$$\frac{C_o + C_1 - (1-p_1)S_1}{p_1} > \frac{C_o + C_1 + (1-p_1)C_c - (1-p_1)(1-p_c)S_c}{p_1 + (1-p_1)p_c}$$

$$\left\{\begin{array}{l} C_op_1 + C_o(1-p_1)p_c + C_1p_1 - (1-p_1)^2S_1p_c \\ +C_1(1-p_1)p_c-(1-p_1)\ S_1p_1 \end{array}\right\} > \left\{\begin{array}{l} C_op_1 + C_1p_1 - (1-p_1)C_cp_1 \\ - (1-p_1)(1-p_c)S_cp_1 \end{array}\right.$$

$$C_1p_c - S_1p_1 + C_op_c - (1-p_1)S_1p_c > C_cp_1 - (1-p_c)S_cp_1$$

$$C_c < \frac{C_op_c - (1-p_1)S_1p_c + (1-p_c)S_cp_1 - S_1p_1 + p_cC_1}{p_1}$$

$$\boxed{C_c < \frac{p_c}{p_1}\left[C_o + C_1 - S_1(1-p_1)\right] + \left[S_c(1-p_c) - S_1\right]}$$

Again for cleanup to be profitable the above equation must hold.

Closed Loop Case

In the simple closed loop case a product is processed through an operation and all parts which pass inspection at the end of the operation go on to the next station, but parts which do not pass inspection are recycled through the operation again. The difference between the closed loop case and the aforementioned recycle case is that under the closed loop case it is assumed that no work has to be done on rejected units before they can be recycled. It is further assumed under the closed loop case that all units which are rejected at the completion of the initial cycle will eventually pass inspection at that station (although some units may require many passes) and therefore, no units will have to be scrapped for a salvage value.

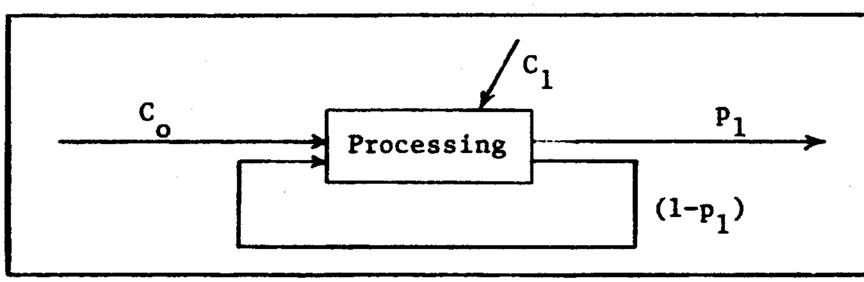


Figure 10

Closed Loop Case

Inspection of Figure 10 will show that the number of units started (N_I) will eventually equal the number of completed good units (N_o).

$$N_I = N_o$$

The total cost per good unit follows:

$$K = \frac{C_o + C_1 - (1-p_1)C_o}{p_1} = C_o + \frac{C_1}{p_1}$$

Recycle with a Primary Loss Case

The recycle with a primary loss case is again a model where spoiled or rejected units are recycled through the process via an expense incurring rework station, but under this case units are recycled due to a rejection following a second station which has received acceptable units from an initial processing station. Upon processing completion at the initial station, units are either accepted and sent on to the second station or rejected and scrapped at a salvage value. At the end of the second station units are either accepted and sent on to the third station or rejected and sent back to the start of the *initial* stage via a recycle station.

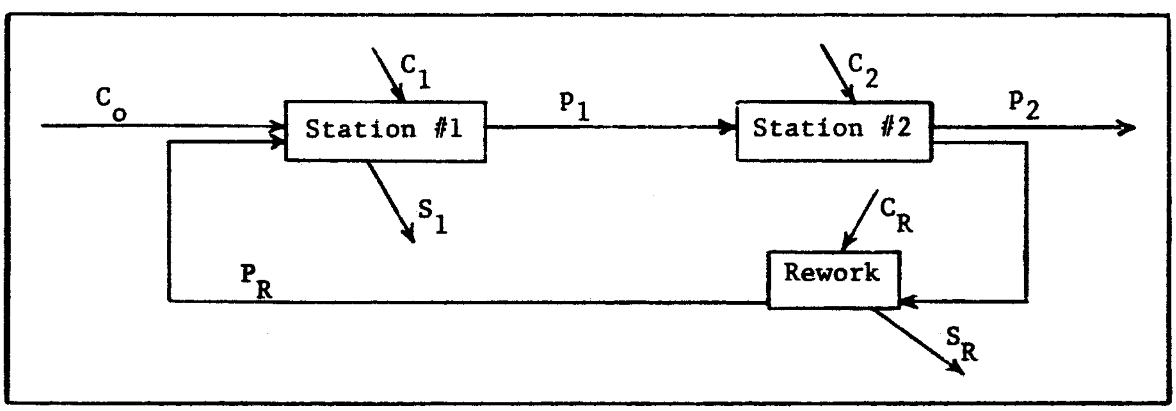


Figure 11

Recycle with a Primary Loss Case

The number of good units through the system:

$$N_o = \frac{p_1 p_2}{1 - p_1 p_R (1-p_2)} N_I$$

The cost of a good unit follows:

$$K = \frac{[C_o+C_1+p_1C_2-(1-p_1)S_1+p_1(1-p_2)C_R-p_1(1-p_2)(1-p_R)S_R-p_1(1-p_2)p_rC_o]N_I}{p_1p_2}$$

<u>Example</u>: C_o = \$1.00; C_1 = \$.20; C_2 = \$.10; C_R = \$.15; S_1 = \$.25; S_R = \$.25

$p_1 = .8$; $p_2 = .9$; $p_R = .7$

<u>Reqd</u>: (a) The number of units which must be started if it is desired to get 100 units out.

(b) Cost of producing 100 good units.

<u>Sol</u>:

(a) $$N_I = \frac{[1 - p_1p_R(1-p_2)]N_o}{p_1p_2}$$

$$= \frac{[1 - (.8)(.7)(1 - .9)]100}{(.8)(.9)} = \frac{(.944)(100)}{.72} = 131 \text{ units}$$

(b) $$K = \frac{131[\$1+.2+(.8)(.1)-(.2)(.25)+(.8)(.1)(.15)-(.8)(.1)(.3)(.25)}{(.8)(.9)}$$

$$\frac{-(.8)(.1)(.7)(1.0)]}{(.8)(.9)} = \frac{(131)(\$1.222)}{.72} = \$222.50$$

Example of Block Combinations

Let's look at an example which requires the use of a combination of the blocks which have been discussed. Notice that products may be extracted from four different locations in the system.

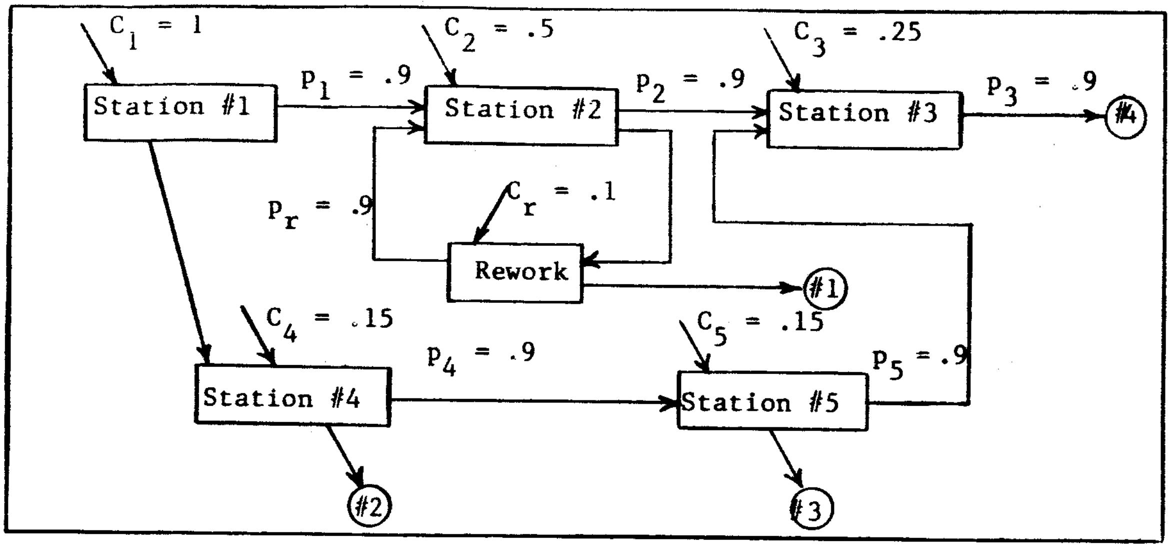


Figure 12

Simulated System Diagram

Reqd: (a) How many units must be started if 1,000 acceptable finished units are needed at location #4?

(b) What is the cost per acceptable finished unit at location #4?

(c) Determine the mathematical models for the output and cost per good unit at locations #1, #2, and #3.

Sol:

(a) First by inspection it can be seen that the probability of a good unit at location #4 is

$$\frac{p_1 p_2 p_3}{1 - (1-p_2)p_r} + (1-p_1)p_4 p_5 p_3 \quad .$$

Then, in order to ascertain the number of units which must be started in order to obtain 1,000 acceptable finished units, we simply divide the required number of finished units by the probability of a single good unit; i.e.,

$$N_o = \left[\frac{p_1 p_2 p_3}{1 - (1-p_2)p_r} + (1-p_1)p_4 p_5 p_3\right] N_I$$

$$1000 = \left[\frac{(.9)(.9)(.9)}{1 - (1-.9)(.9)} + (1-.9)(.9)(.9)(.9) \right] N_I$$

$$1000 = .8729\ N_I$$

or

$$N_I = 1146 \text{ units}$$

(b) Likewise, the cost per good unit at location #4 can also be determined by inspection; i.e.,

$$K = \frac{C_1 + \dfrac{p_1[C_2 + (1-p_2)C_r]}{1 - (1-p_2)p_r} + \dfrac{p_1 p_2 C_3}{1 - (1-p_2)p_r} + (1-p_1)C_4 + (1-p_1)p_4 C_5 + (1-p_1)p_4 p_5 C_3}{\dfrac{p_1 p_2 p_3}{1 - (1-p_2)p_r} + (1-p_1)p_4 p_5 p_3}$$

$$K = \frac{1 + \dfrac{(.9)[.5+(.1)(.1)]}{1 - (.1)(.9)} + \dfrac{(.9)(.9)(.25)}{1 - (.1)(.9)} + (.1)(.15)+(.1)(.9)(.15) + (.1)(.9)(.9)(.25)}{\dfrac{(.9)(.9)(.9)}{1-(.1)(.9)} + (.1)(.9)(.9)(.9)}$$

$$K = \frac{1 + .505 + .222 + .0487}{.8729}$$

$$K = \$2.03 \text{ per unit}$$

Thus it is shown that for the costs and efficiencies depicted in the above illustration, 1,146 units must be started to obtain 1,000 good units at a cost of $2.03 per unit — $2,030 total cost.

(c) By inspection, the mathematical models for the required number of units to be started (N_I) to get a given number of units out is as follows:

$$N_o = \frac{p_1(1-p_2)(1-p_r)}{1 - (1-p_2)p_r} N_I$$

Likewise, the cost per good finished unit is as follows:

$$K = \frac{C_1 + \dfrac{p_1[C_2 + (1-p_2)C_r]}{1 - (1-p_2)p_r}}{\dfrac{p_1(1-p_2)(1-p_r)}{1 - (1-p_2)p_r}} = \frac{C_1[1 - (1-p_2)p_r] + p_1[C_2 + (1-p_2)C_r]}{p_1(1-p_2)(1-p_r)}$$

The mathematical models for units started and costs at locations #2 and #3 are found the same way and are as follows:

At location #2:

$$N_o = [(1-p_1)(1-p_4)]N_I$$

$$K = \frac{C_1 + (1-p_1)C_4}{[(1-p_1)\ (1-p_4)]}$$

At location #3:

$$N_o = [(1-p_1)p_4(1-p_5)]N_I$$

$$K = \frac{C_1 + (1-p_1)C_4 + (1-p_1)p_4C_5}{[(1-p_1)p_4(1-p_5)]}$$

It may be interesting now to look at an actual production system and to see how the concept of cost models can be applied. This system, which consists of the assembly of semiconductors, is relatively complex and will serve to show that the application of cost models is feasible regardless of the complexity of the system. The diagram below shows the flow of the assembly from station to station with the probability of a good unit out of a station. The associated labor cost at a station is shown as L-xxxx and the material cost is shown as M-xxxx. A numerical solution for the system output and cost per unit is not presented, but could easily be calculated using the techniques described previously.

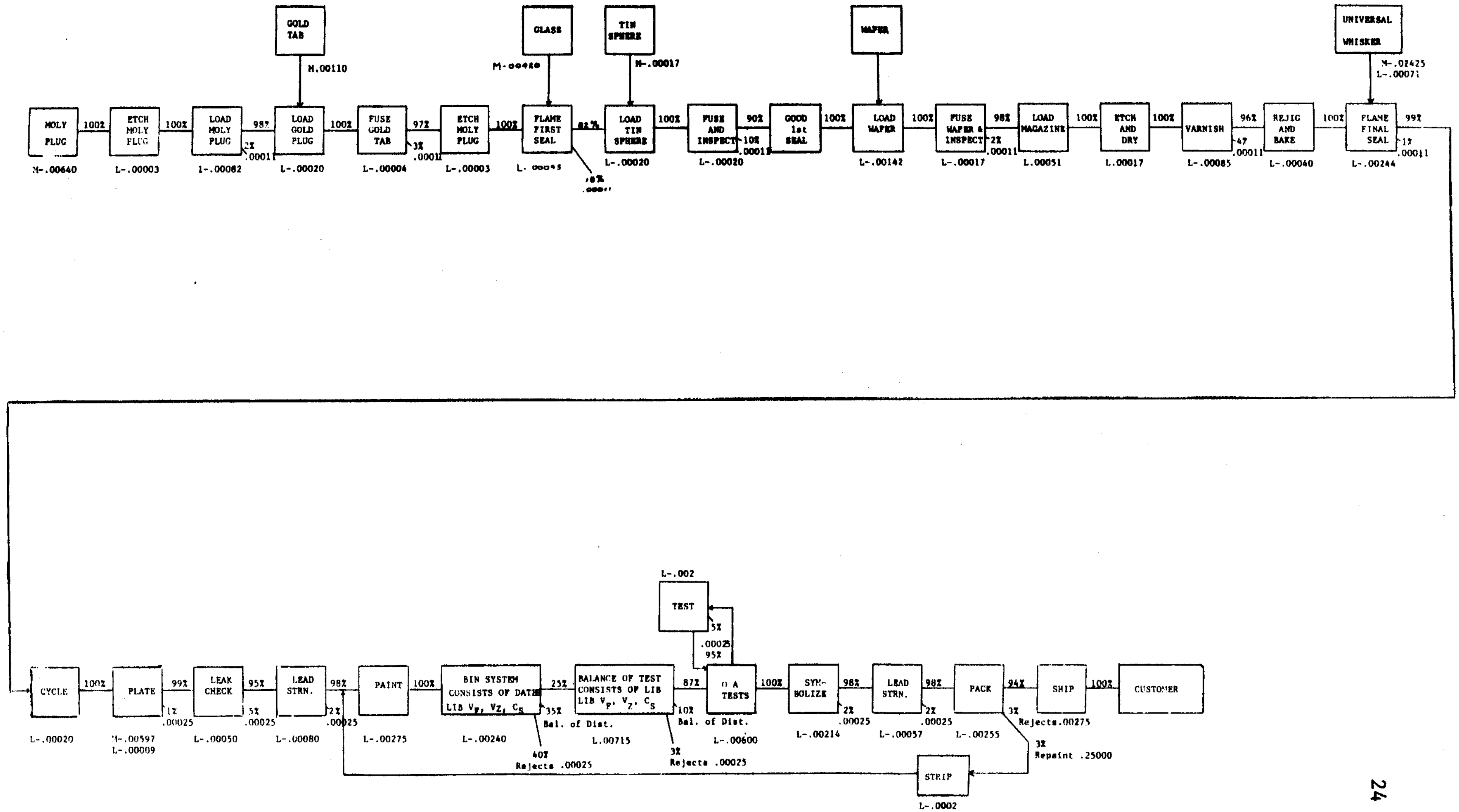


Figure 13

Semiconductor Production System

III. OPTIMIZING PROFIT ON A MULTIPRODUCT SYSTEM

Most of our discussion in this paper so far has been restricted to the determination of system efficiencies and cost per unit of product. Now let's expand these ideas and apply them to a system which produces several different products at different locations along the production line. This situation could occur when, after processing at a station, some units are drained off and sold, while the remaining units are sent to the next processing station to be made into other types of products. This system is used when there is a known market for each product and the problem that the management must answer is how many raw units must be started initially in order to maximize the profit on the multiproduct system. Under these conditions it can be seen that profits on the products can be optimized, but the total cost of production will not necessarily be optimized; i.e., it may be less expensive to produce at volumes other than the volume which optimizes profit.

This type of problem could be solved by the use of linear or integer programming, but it will be shown that the problem can also be solved using the feedback block diagram theory which has been discussed in this paper. The example below will illustrate the applicability of this method on a typical production system.

Example:

In the system below assume that there are four different products extracted from four different locations along the production line. Assume that we know the consumer demand (N) for each product, the price per unit (P) which will be charged for each product, and the marketing expense per unit (M) associated with each product. Further assume that any overproduction on any product is worthless.

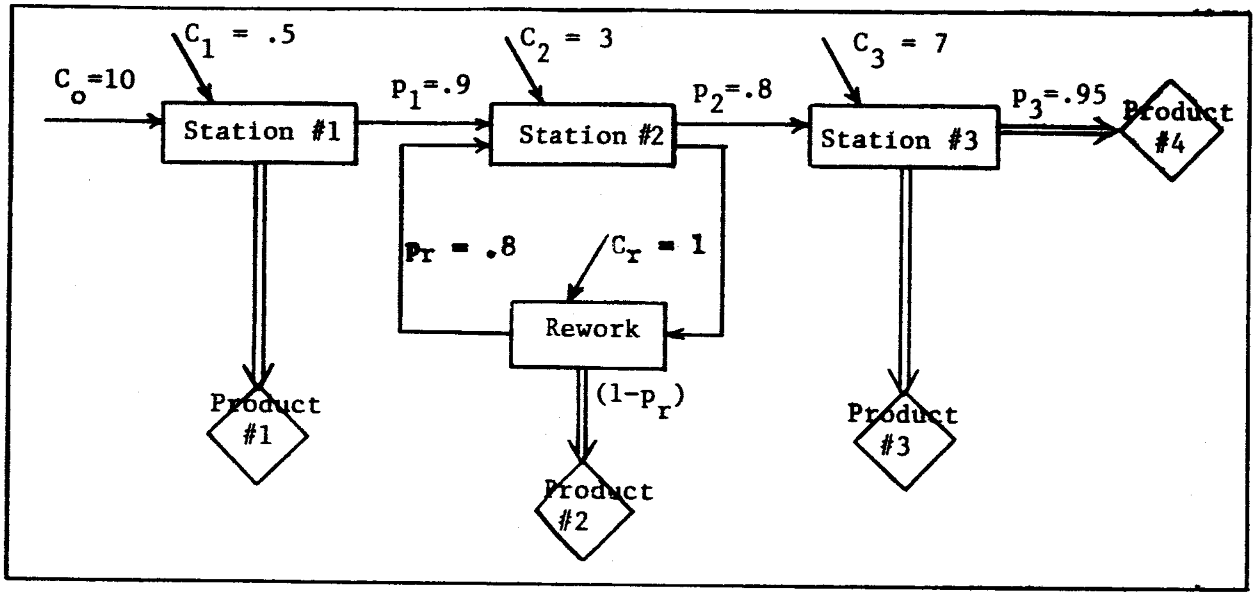


Figure 14

Simulated Multiproduct System

Let $N_1 = 1000$ $\quad M_1 = \$1.00$ $\quad P_1 = \$100$

$N_2 = 1500$ $\quad M_2 = \$4.00$ $\quad P_2 = \$200$

$N_3 = 2000$ $\quad M_3 = \$3.00$ $\quad P_3 = \$250$

$N_4 = 300$ $\quad M_4 = \$2.00$ $\quad P_4 = \$500$

N_o = Number of units started

First we calculate the number of units which must be started to satisfy the demand for each product.

$$N_{o_1} = N_1/(1-p_1) = 1000/(1-.9) = 10{,}000 \text{ units}$$

$$N_{o_2} = \frac{[N_2/p_1(1-p_2)(1-p_r)]}{1-(1-p_2)p_r} = \frac{[1500/(.9)(.2)(.2)]}{1-(.2)(.8)} = 34{,}965 \text{ units}$$

$$N_{o_3} = \frac{[N_3/p_1 p_2 (1-p_3)]}{1 - (1-p_2)p_r} = \frac{[2000/(.9)(.8)(.05)]}{1 - (.2)(.8)} = 46{,}620 \text{ units}$$

$$N_{o_4} = \frac{N_4/p_1 p_2 p_3}{1-(1-p_2)p_r} = \frac{[300/(.9)(.8)(.95)]}{1 - (.2)(.8)} = 368 \text{ units}$$

Next we calculate the manufacturing cost through the system.

$$\text{Mfg. cost} = N_o \left\{ (C_o + C_1) + p_1 \left[\frac{C_2 + (1-p_2)C_r}{1 - (1-p_2)\ p_r} \right] + \frac{p_1 p_2 C_3}{1-(1-p_2)p_r} \right\}$$

Then the equation for total profit follows:

$$\text{Profit} = \sum_{i=1}^{4} N_i P_i - \sum_{i=1}^{4} N_i M_i - \text{mfg. cost}$$

Since profit is a linear function of N, we only need to check the profit at the four volume levels indicated which will satisfy the four required demands respectively.

<u>Check profit at N_o = 368 units:</u>

$$\text{Mfg. cost} = 368 \left\{ (15) + (.9) \left[\frac{3 + (.2)(.1)}{1 - (.2)(.8)} \right] + \frac{(.9)(.8)(.7)}{1 - (.2)(.8)} \right\}$$

$$= (368)(\$24.43) = \$8{,}990$$

$$\text{Profit} = \sum_{i=1}^{4} P_i N_i - \sum_{i=1}^{4} M_i N_i - \text{mfg. cost}$$

$$P_1 N_1 = (N_o)(1-p_1)\ (P_1) = (368)(.1)(\$100) = (36)(\$100) = \$3{,}600$$

$$P_2 N_2 = N_o \left[\frac{p_1 (1-p_2)(1-p_r)}{1 - (1-p_2)p_r} \right] (P_2) = (368) \left[\frac{(.9)(.2)(.2)}{1 - (.2)(.8)} \right] (\$200)$$

$$= (368)(.0429) \times (\$200) = (15)(\$200) = \$3{,}000$$

$$P_3N_3 = N_o \left[\frac{p_1p_2(1-p_3)}{1-(1-p_2)p_r}\right] (p_3) = (368)\left[\frac{(.9)(.8)(.05)}{.84}\right] (\$250)$$

$$= (368)(.0429) \times (\$250) = (15)(\$250) = \$3,750$$

$$P_4N_4 = N_o \left[\frac{p_1p_2p_3}{1-(1-p_2)p_r}\right] \times P_4 = (368)\left[\frac{(.9)(.8)(.95)}{.84}\right] (\$500)$$

$$= (368)(.814) \times (\$500) = (299)(\$500) = \$149,500$$

$$\text{Profit} = \$3600 + \$3000 + \$3750 + \$149,500 - (36)(\$1) - (15)(\$4)$$

$$- (15)(\$3) - (299)(\$2) - \$8990$$

$$= \$159,850 - \$739 - \$8990$$

$$= \$150,121$$

<u>Check profit at N_o = 10,000 units:</u>

$$\text{Mfg. cost} = 10,000(\$24.43) = \$244,300$$

$$P_1N_1 = (10,000/.1)(\$100) = \$100,000$$

$$P_2N_2 = (10,000)(.0429) \times (\$200) = (429)(\$200) = \$85,800$$

$$P_3N_3 = (10,000)(.0429) \times (\$250) = (429)(\$250) = \$107,250$$

$$P_4N_4 = (10,000)(.814) \times (\$500) = (8140)(\$500) = (300)^*(\$500) = \$150,000$$

$$\text{Profit} = \$100,000 + \$85,800 + \$107,250 + \$150,000 - (1000)(\$1)$$

$$- (429)(\$4) - (429)(\$3) - (300)(\$2) - \$244,300$$

$$= \$433,050 - \$4603 - \$244,300$$

$$= \$194,147$$

*8140 units exceed the demand for produt #4

<u>Check profit at N_o = 34,965 units:</u>

Mfg. cost = 34,965($24.43) = $854,195

P_1N_1 = (34,965)(.1)($100) = (3496)($100) = (1000)($100) = $100,000

P_2N_2 = (34,965)(.0429) × ($200) = (1500)($200)

= (1500)($200) = $300,000

P_3N_3 = (34,965)(.0429) × ($250) = (1500)($250)

= (1500)($250) = $375,000

P_4N_4 = (34,965)(.814) × ($500) = (28,462)($500) = (300)($500)

= $150,000

Profit = $100,000 + $300,000 + $375,000 + $150,000 - (1000)($1)

- (1500)($4) - (1500)($3) - (300)($2) - $854,195

= $925,000 - $12,100 - $854,195

= $58,706

<u>Check profit at N_o = 46,620 units:</u>

Mfg. cost = 46,620($24.43) = $1,138,927

P_1N_1 = (1000)($100) = $100,000

P_2N_2 = (1500)($200) = $300,000

P_3N_3 = (2000)($250) = $500,000

P_4N_4 = (300)($500) = $150,000

Profit = $100,000 + $300,000 + $500,000 + $150,000

-(1000)($1) - (1500)($4) - (2000)($3) - (300)($2) - $1,138,927

= $1,050,000 - $13,600 - $1,138,927

= -$102,527

A graphical plot of the number of units started versus the total profit shows that the optimum production quantity is 10,000 units.

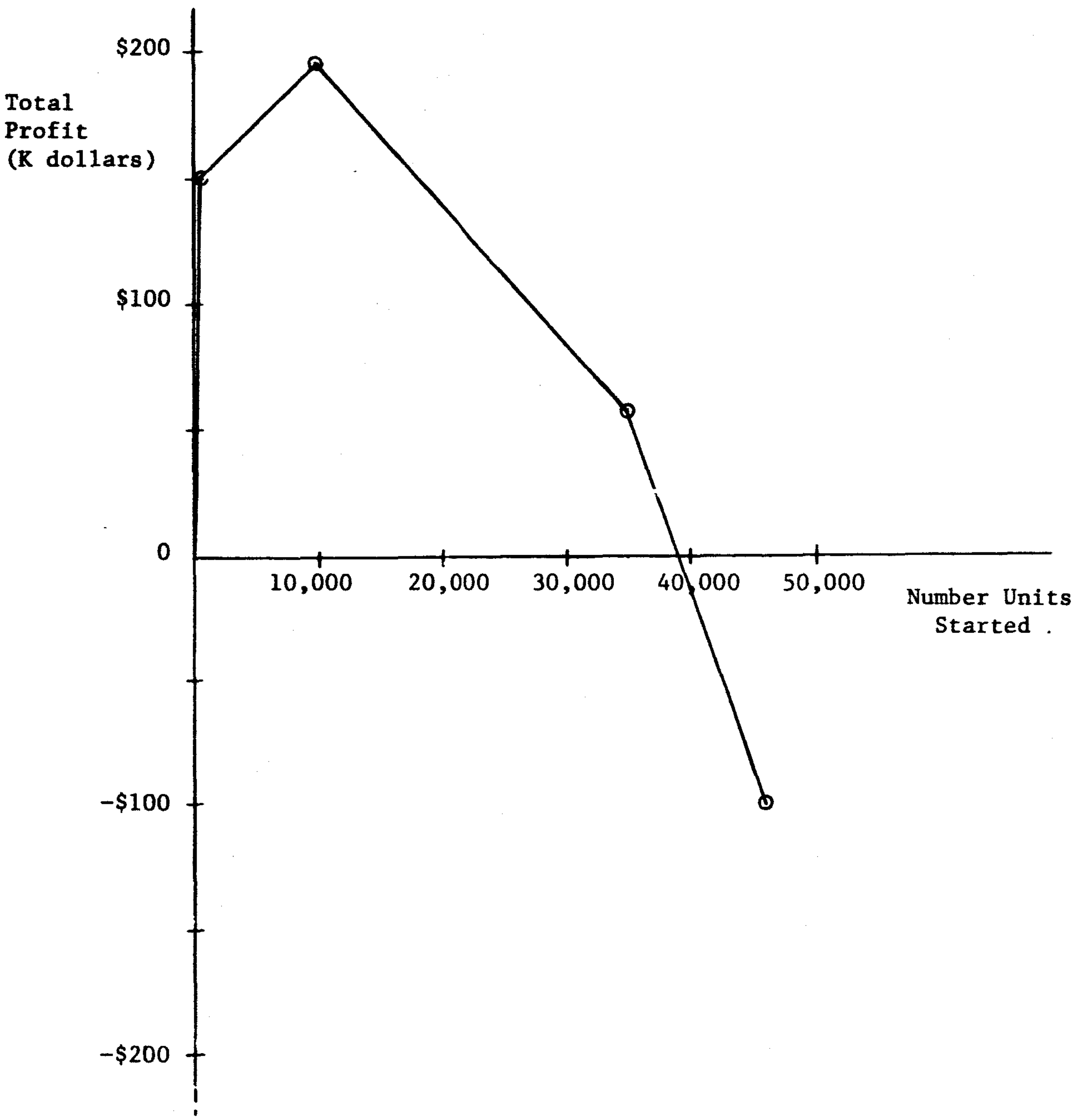


Figure 15

Graph of Total Profit vs.
Number of Units Started

IV. SUMMARY

This paper has presented the use of the transfer function block diagram method to investigate elementary manufacturing problems. It is believed that this method has many advantages over the standard cost system method of evaluating production costs and efficiencies. The most obvious of these is that it gives a much more accurate indication of the true cost of a process. There are many instances when management has chosen one course of action, given the result of a standard cost analysis when, in reality, another course would have been optimum. We believe that the use of production cost models will most often give this optimum result. The methods application to these other problems is left to the reader.

BIBLIOGRAPHY

Articles

Brabston, Donald C., "Understanding and Utilization of Standard Costs," *The Journal of Accountancy*, May, 1959; p. 52.

Wortham, Albert W., "Servo Theory in Costing," unpublished notes, Texas A&M University, College Station, Texas

Books

Barnes, Ralph M., *Motion and Time Study*, John Wiley & Sons, Inc., New York, 1940.

Browning, Donald B., *Direct Costing*, A Thesis, January, 1960, pp. 10-24.

Camman, Eric A., *Basic Standard Costs*, American Institute Publishing Company, Inc., New York, 1932.

Henrici, Stanley B., *Standard Costs for Manufacturing*, McGraw-Hill Book Company, New York, January, 1947.

Khan, Nurul A., *Standard Costs for Labor*, A Thesis, January, 1960, pp. 3-8.

Niebel, Benjamin W., *Motion and Time Study*, Richard D. Irwin, Inc., Homewood, Illinois, April, 1955.

Pamphlets

Flewellen, William C., "An analysis of Cost Control Problems of Small Manufacturers in Alabama," Research Reports, University of Alabama, January, 1961.

Goodey, F. R., "A Practical Application of Standard Costs," Mechanical World Monograph #43, Emmott & Co., Ltd., London, 1948.

APPENDIX A

DETERMINATION OF THE EXPECTED VALUE AND VARIANCE OF THE COST FOR THE SIMPLE BLOCK MODEL

This appendix presents a proof that the cost per good unit through a single work station (K_s) is actually $(C_o + C_1)/p_1$ (assuming zero salvage value for simplicity).

First, let x = 1 when a good unit is produced and x = 0 when a bad unit is produced. Then,

$$E(K) = \int Kf(K)dK$$

$$f(y) = (1-p)^{y-1}p, \text{ where } y = 1, 2, \ldots$$

which is the probability of one good unit in y trials.

$$E(K) = E[(C_o + C_1)y] = (C_o + C_1)\ E(y)$$

$$E(y) = \sum_{y=1}^{\infty} yf(y) = \sum_{y=1}^{\infty} y(1-p)^{y-1}p$$

Recall

$$g(p) = \sum_{y=1}^{\infty} (1-p)^{y} = (1-p) \sum_{y=1}^{\infty} (1-p)^{y-1}$$

$$= (1-p) \sum_{y=0}^{\infty} (1-p)^{y} = (1-p) \frac{1}{1 - (1-p)} = \frac{1-p}{p}$$

Then

$$\frac{dg(p)}{dp} = - \sum_{y=1}^{\infty} y(1-p)^{y-1} = \frac{d[(1-p)/p]}{dp}$$

Therefore,

$$E(y) = \sum_{y=1}^{\infty} (1-p)^{y-1}p = \frac{-pd[(1-p)/p]}{dp} = (-p)[1/p(-1) + (1-p)(-1/p^2)]$$

$$= (-p)[-1/p - (1-p)/p^2] = (-p)[-p - (1-p)/p^2] = (-p)(-1/p^2)$$

$$= 1/p$$

$$E(K) \equiv (C_o + C_1)/p$$

To evaluate the variance of K_s:

$$V(K) = E[(k - K_{avg})^2] = E(K^2) - (K_{avg})^2$$

$$= E[(C_o + C_1)y]^2 - [(C_o + C_1)/p]^2 = (C_o + C_1)^2 E(y^2) - (C_o + C_1)^2/p^2$$

Now we need to know the expected value of y^2.

$$E(y^2) = \sum_{y=1}^{\infty} y^2 (1-p)^{y-1} p$$

$$\frac{dg(p)}{dp} = \frac{d[(1-p)/p]}{dp} = -\frac{1}{p^2}$$

$$\frac{d^2 g(p)}{dp^2} = \sum_{y=1}^{\infty} (y)(y-1)(1-p)^{y-2} = \frac{2}{p^3}$$

Now multiply out the terms

$$\sum_{y=1}^{\infty} y^2 (1-p)^{y-2} - \sum_{y=1}^{\infty} y(1-p)^{y-2} = \frac{2}{p^3}$$

$$\frac{1}{1-p} \sum_{y=1}^{\infty} y^2 (1-p)^{y-1} - \frac{1}{1-p} \sum_{y=1}^{\infty} y(1-p)^{y-1} = \frac{2}{p^3}$$

$$\frac{1}{1-p} \sum_{y=1}^{\infty} y^2 (1-p)^{y-1} - \frac{1}{p^2(1-p)} = \frac{2}{p^3}$$

$$\sum_{y=1}^{\infty} y^2 (1-p)^{y-1} = \left[\frac{2}{p^3} + \frac{1}{p^2(1-p)}\right] (1-p)$$

Therefore,

$$E(y^2) = p\left[\frac{2}{p^3} + \frac{1}{p^2(1-p)}\right](1-p) = \left[\frac{2(1-p) + p}{p^2(1-p)}\right] (1-p)$$

$$E(y^2) = \frac{2 - 2p + p}{p^2} = \frac{2 - p}{p^2}$$

Then

$$V(K) = (C_o + C_1)^2 \, [(2-p)/p^2 - 1/p^2] = (C_o + C_1)^2 \, [(1-p)/p^2]$$

or

$$\sigma_K = (C_o + C_1) \sqrt{1-p}/p$$

APPENDIX B

DETERMINATION OF THE EXPECTED VALUE AND VARIANCE OF THE COST FOR THE SIMPLE BLOCKS IN SERIES MODEL

In Appendix A it was shown how to calculate the expected value and variance of the cost per good unit through a single work station. In this appendix it will be shown that, although slightly more tedious, the expected value and variance of the cost through a multistage system can be calculated with the same approach. As an example we will show that the cost through three stages (K_3) is equal to

$$\frac{C_o + C_1}{p_1 p_2 p_3} + \frac{C_2}{p_2 p_3} + \frac{C_3}{p_3} \quad .$$

As shown previously, $E(K_1) = (C_o + C_1)/p_1$. Then,

$$E(K_2) = E\left[\left(\frac{C_o + C_1}{p_1} + C_2\right)y_2\right] = \left[\frac{C_o + C_1}{p_1} + C_2\right] E(y_2)$$

$$E(y_2) = \sum_{y=1}^{\infty} y_2 f(y_2) = \sum_{y=1}^{\infty} y_2 (1-p_2)^{y_2 - 1} p_2$$

As before, substitution yields

$$E(y_2) = \frac{-p_2 d[(1-p_2)/p_2]}{dp_2} = (-p_2)[1/p_2(-1) + (1-p_2)(-1/p_2)]$$

$$= (-p_2)[-1/p_2 - (1-p_2)/p_2^2\,] = (-p_2)[-p_2 - (1-p_2)/p_2^2\,]$$

$$= (-p_2)(-1/p_2^2) = 1/p_2$$

Therefore,

$$E(K_2) = \left[\frac{C_o + C_1}{p_1} + C_2\right] \frac{1}{p_2} = \frac{(C_o + C_1)/p_1 + C_2}{p_2}$$

$$= \frac{C_o + C_1}{p_1 p_2} + \frac{C_2}{p_2}$$

Then,

$$E(K_3) = E\left[\frac{C_o + C_1}{p_1 p_2} + \frac{C_2}{p_2} + C_3 \quad y_3\right] = \left[\frac{C_o + C_1}{p_1 p_2} + \frac{C_2}{p_2} + C_3\right] E(y_3)$$

$$E(y_3) = \sum_{y=1}^{\infty} y_3 f(y_3) = \sum_{y=1}^{\infty} y_3 (1-p_3)^{y_3 - 1} p_3$$

Substitution yields

$$E(y_3) = 1/p_3$$

Therefore,

$$E(K_3) = \left[\frac{C_o + C_1}{p_1 p_2} + \frac{C_2}{p_2} + C_3\right] \frac{1}{p_3}$$

$$= \frac{[(C_o + C_1)/p_1 p_2] + (C_2/p_2) + C_3}{p_3} \equiv \frac{C_o + C_1}{p_1 p_2 p_3} + \frac{C_2}{p_2 p_3} + \frac{C_3}{p_3}$$

To evaluate the variance of $\{(((C_o + C_1)y_1 + C_2)y_2 + C_3)y_3\}$

$$V\{(((C_o + C_1)y_1 + C_2)y_2 + C_3)y_3\} = E\{[(((C_o + C_1)y_1 + C_2)y_2 + C_3)y_3]^2 - \left[\frac{C_o + C_1}{p_1p_2p_3} + \frac{C_2}{p_2p_3} + \frac{C_3}{p_3}\right]^2\}$$

$$= E[(((C_o + C_1)y_1 + C_2)y_2 + C_3)y_3]^2 - \left[\frac{C_o+C_1}{p_1p_2p_3} + \frac{C_2}{p_2p_3} + \frac{C_3}{p_3}\right]^2$$

$$= E[(C_o + C_1)^2y_1^2y_2^2y_3^2] + E[C_2^2y_2^2y_3^2] + E[C_3^2y_3^2]$$

$$+ E[2(C_o + C_1)y_1y_2y_3C_2y_2y_3] + E[2(C_o + C_1)y_1y_2y_3C_3y_3]$$

$$+ E[2C_2y_2y_3C_3y_3] - \frac{(C_o + C_1)^2}{p_1^2p_2^2p_3^2} + \frac{C_2^2}{p_2^2p_3^2} - \frac{C_3^2}{p_3^2}$$

$$- \frac{2(C_o + C_1)(C_2)}{p_1p_2^2p_3^2} - \frac{2(C_o + C_1)(C_3)}{p_1p_2p_3^2} - \frac{2C_2C_3}{p_2p_3^2}$$

$$= (C_o+C_1)^2 \frac{(2-p_1)^2}{p_1^4}\frac{(2-p_2)^2}{p_2^4}\frac{(2-p_3)^2}{p_3^4} + C_2^2\frac{(2-p_2)^2}{p_2^4}\frac{(2-p_3)^2}{p_3^4} + C_3^2\frac{(2-p_3)^2}{p_3^4}$$

$$+ 2(C_o+C_1)C_2\frac{(2-p_1)}{p_1^2}\frac{(2-p_2)^2}{p_2^4}\frac{(2-p_3)^2}{p_3^4} + 2(C_o+C_1)C_3\frac{(2-p_1)}{p_1^2}\frac{(2-p_2)}{p_2^2}\frac{(2-p_3)^2}{p_3^4}$$

$$+ 2C_2C_3\frac{(2-p_2)}{p_2^2}\frac{(2-p_3)^2}{p_3^4} - \frac{(C_o+C_1)^2}{p_1^2p_2^2p_3^2} - \frac{C_2^2}{p_2^2p_3^2} - \frac{C_3^2}{p_3^2} - \frac{2(C_o+C_1)C_2}{p_1p_2^2p_3^2}$$

$$\frac{-2(C_o+C_1)C_3}{p_1p_2p_3^2} - \frac{2C_2C_3}{p_2p_3^2}$$

$$= \frac{(C_o+C_1)^2}{p_1^2p_2^2p_3^2}\left[\frac{(2-p_1)^2}{p_1^2}\frac{(2-p_2)^2}{p_2^2}\frac{(2-p_3)^2}{p_3^2} - 1\right] + \frac{C_2^2}{p_2^2p_3^2}\left[\frac{(2-p_2)^2}{p_2^2}\frac{(2-p_3)^2}{p_3^2} - 1\right]$$

$$+ \frac{C_3^2}{p_3^2}\left[\frac{(2-p_3)^2}{p_3^2} - 1\right] + \frac{2(C_o+C_1)C_2}{p_1p_2^2p_3^2}\left[\frac{(2-p_1)(2-p_2)^2\ (2-p_3)^2}{p_1p_2^2p_3^2} - 1\right]$$

$$+ \frac{2(C_o+C_1)C_3}{p_1p_2p_3^2}\left[\frac{(2-p_1)(2-p_2)(2-p_3)^2}{p_1p_2p_3^2} - 1\right] + \frac{2C_2C_3}{p_2p_3^2}\left[\frac{(2-p_2)(2-p_3)^2}{p_2p_3^2} - 1\right]$$